

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2003, 11:03:24 ; Search time 41 Seconds
(without alignments)
963.974 Million cell updates/sec

Title: US-09-987-755-2
Perfect score: 1319
Sequence: 1 MAAVSLRLGDLVWGKLGKRYP.....VCYQAITKKLKICEDLLLP 249

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1319	100.0	249	AAW09404	Human hepatoma-der
2	1319	100.0	260	AAU20553	Human secreted pro
3	1319	100.0	260	AAU21747	Novel human neopla
4	1314	99.6	250	AAU21590	Novel human neopla
5	1291	97.9	547	AAW69240	Clone AQ73_3 prote
6	1291	97.9	547	AAW90720	Human AQ73_3 prot
7	1291	97.9	553	AAO16415	Human nucleic acid
8	1278	96.9	550	AAU21690	Novel human neopla
9	1263	95.8	269	AAU20515	Human secreted pro

10	1263	95.8	269	22	AAU21813	Novel human neopla
11	193	14.6	475	22	ABB58603	Drosophila melanog
12	190.5	14.4	202	23	ABB72327	Rat protein isolat
13	189	14.3	203	20	AAU28502	HGFH2 Human Growth
14	189	14.3	203	21	AAU13521	Human hepatoma-der
15	189	14.3	203	21	AAU88374	Human type II hepa
16	189	14.3	203	22	AAU92800	Human protein sequ
17	189	14.3	203	22	AAU70073	Human secreted pro
18	189	14.3	203	23	ABG65515	Human albumin fusi
19	186.5	14.1	246	21	AAU58173	Lung cancer associ
20	185.5	14.1	237	18	AAU33216	Mouse hepatoma der
21	185.5	14.1	240	16	AAU66727	Human hepatoma der
22	185.5	14.1	240	21	AAU09984	Human HDGFIP prote
23	185.5	14.1	240	22	ABB50263	Hepatoma-derived g
24	178.5	13.5	235	19	AAU61963	Lung growth factor
25	166.5	12.6	326	22	ABB12116	Human PRO1604 homo
26	165	12.5	314	22	AAU70075	Human secreted pro
27	165	12.5	314	23	ABG65517	Human albumin fusi
28	165	12.5	671	21	AAU99426	Human PRO1604 (UNQ
29	165	12.5	671	22	AAU29196	Human PRO polypept
30	165	12.5	671	22	AAU66175	Protein of the inv
31	165	12.5	671	23	AAU97038	Human LP protein L
32	165	12.5	671	23	ABB05042	Human NOV4d protei
33	165	12.5	671	24	ABU71284	Human PRO1604 prot
34	165	12.5	671	24	ABU65741	Human secreted/tra
35	165	12.5	671	24	ABU66074	Novel human secret
36	165	12.5	671	24	ABU67578	Human secreted/tra
37	165	12.5	671	24	ABU65436	Human PRO polypept
38	165	12.5	671	24	ABU58572	Human PRO polypept
39	165	12.5	671	24	ABU56108	Human secreted/tra
40	165	12.5	671	24	ABU57103	Human PRO polypept
41	165	12.5	671	24	ABU10682	Human secreted/tra
42	165	12.5	676	22	AAU47218	Human NOV9 protein
43	165	12.5	676	22	AAU64373	Human hepatome cel
44	164	12.4	667	23	ABB05040	Human NOV4b protei
45	164	12.4	667	23	ABB05041	Human NOV4c protei

ALIGNMENTS

RESULT 1
AAW09404
ID AAW09404 standard; Protein; 249 AA.
XX
AC AAW09404;
XX
DT 20-MAY-1997 (first entry)
XX
DE Human hepatoma-derived growth factor (HDGF-2).
XX
KW Hepatoma-derived growth factor-2; HDGF-2; chemokine; wound healing;
KW vulnery; burn; ulcer; thrombosis; arteriosclerosis; antibody;
KW antagonist; tumour; therapy.
XX
OS Homo sapiens.
XX
PN WO9639485-A1.
XX
PD 12-DEC-1996.
XX
PF 05-JUN-1995; 95WO-US06731.
XX
PR 05-JUN-1995; 95WO-US06731.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Kunsch CA, Rosen CA;
XX
DR WPI; 1997-043108/04.
DR N-PSDB; AAT47520.
XX
PT DNA encoding human-derived growth factor polypeptide - useful to

PT promote wound healing as result of burns or ulcers
XX
XX Claim 7; Fig 2; 54pp; English.
XX Human hepatoma-derived growth factor-2 (HDGF-2) (AAW09404) shows
CC 23% identity and 61% similarity over a 201-amino acid stretch
CC to human HDGF, a mitogen for fibroblasts. Its amino acid sequence
CC was deduced from a cDNA clone (AA747520) isolated from a human
CC umbilical vein endothelial tissue cDNA library. Recombinant
CC HDGF-2 can be produced in transformed host (e.g. E. coli, Sf9, COS)
CC cells. It is useful in stimulating tissue repair and tissue growth
CC and can also be used to produce antibodies and to screen for
CC antagonists that retard tumour growth and reduce scarring.
XX
SQ Sequence 249 AA;

Query Match 100.0%; Score 1319; DB 18; Length 249;
Best Local Similarity 100.0%; Pred. No. 1.3e-121;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAVSLRLGDLVWGKLGYPWPVGKIVNPPKDLKPRGKKCFVFFGTEHAWIKVEQL 60
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1 MAAVSLRLGDLVWGKLGYPWPVGKIVNPPKDLKPRGKKCFVFFGTEHAWIKVEQL 60

QY 61 KPYHAHKEEMIKKNGKRFQQAQVDAVEEFLRRAGKDKQTSSHNSDDKNNRNSSEERSRP 120
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
61 KPYHAHKEEMIKKNGKRFQQAQVDAVEEFLRRAGKDKQTSSHNSDDKNNRNSSEERSRP 120

QY 121 NSGDEKRLKLSLSEGVKKNMGEGKKRVSSGSSERGSKSPKRAQEQSPKRGPPKDEKD 180
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
121 NSGDEKRLKLSLSEGVKKNMGEGKKRVSSGSSERGSKSPKRAQEQSPKRGPPKDEKD 180

QY 181 LTIPESSTVKGMMAGPMAAFKWQPTASEPVKDADPHFHHLLSQTEKPAVCYQAITKKLK 240
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
181 LTIPESSTVKGMMAGPMAAFKWQPTASEPVKDADPHFHHLLSQTEKPAVCYQAITKKLK 240

QY 241 ICEDLLLP 249
Db ||||||||
241 ICEDLLLP 249

RESULT 2
AAU20553
ID AAU20553 standard; Protein; 260 AA.
XX
AC AAU20553;
XX
DT 04-DEC-2001 (first entry)
DE Human secreted protein, Seq ID No 545.
XX
KW Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
KW rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW Gaucher's disease; neurological disease; cerebrovascular disorder;
KW thrombosis; wound healing.
XX
OS Homo sapiens.
XX
PN WO20015326-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01347.
XX
PR 31-JAN-2000; 2000US-0179065.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX

DR WPI; 2001-451931/48.
DR N-PSDB; AAS33262.
XX
PT New nucleic acids and polypeptides, useful for diagnosing, preventing
PT or treating medical conditions -
XX
PS Claim 11; SEQ ID No 545; 753pp; English.
XX
CC The invention relates to novel isolated nucleic acid molecules (I)
CC encoding human secreted proteins (II). (I) and (II) are used to prevent,
CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
CC the prevention, treatment and diagnosis of diseases associated with
CC inappropriate expression of secreted proteins. (I) and complementary
CC sequences may also be used as DNA probes in diagnostic assays (e.g.
CC polymerase chain reactions (PCR)) to detect and quantitate the presence
CC of similar nucleic acid sequences in samples, and so which patients may
CC be in need of restorative therapy. (II) may also be used as antigens in
CC the production of antibodies and in assays to identify modulators
CC (agonists and antagonists) of the expression and activity of the secreted
CC proteins. The anti-(II) antibodies and antagonists may also be used to
CC down regulate expression and activity of (II). The anti-(II) antibodies
CC may also be used as diagnostic agents for detecting the presence of (II)
CC in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
CC disorders include for example: immune/autoimmune diseases (e.g. HIV
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
CC angina and thrombosis), infections caused by bacteria, viruses and
CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),
CC agonists, antagonists and antibodies can also be used to promote wound
CC healing, maintain organs before transplantation, and support cell culture
CC of primary tissues. AAU20342-AAU20666 represent human secreted protein
CC amino acid sequences, and related sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification but was obtained in electronic format directly from WIPO
CC at: ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 260 AA;

Query Match 100.0%; Score 1319; DB 22; Length 260;
Best Local Similarity 100.0%; Pred. No. 1.4e-121;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAVSLRLGDLVWGKLGYPWPVGKIVNPPKDLKPRGKKCFVFFGTEHAWIKVEQL 60
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
12 MAAVSLRLGDLVWGKLGYPWPVGKIVNPPKDLKPRGKKCFVFFGTEHAWIKVEQL 71

QY 61 KPYHAHKEEMIKKNGKRFQQAQVDAVEEFLRRAGKDKQTSSHNSDDKNNRNSSEERSRP 120
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
72 KPYHAHKEEMIKKNGKRFQQAQVDAVEEFLRRAGKDKQTSSHNSDDKNNRNSSEERSRP 131

QY 121 NSGDEKRLKLSLSEGVKKNMGEGKKRVSSGSSERGSKSPKRAQEQSPKRGPPKDEKD 180
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
132 NSGDEKRLKLSLSEGVKKNMGEGKKRVSSGSSERGSKSPKRAQEQSPKRGPPKDEKD 191

QY 181 LTIPESSTVKGMMAGPMAAFKWQPTASEPVKDADPHFHHLLSQTEKPAVCYQAITKKLK 240
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
192 LTIPESSTVKGMMAGPMAAFKWQPTASEPVKDADPHFHHLLSQTEKPAVCYQAITKKLK 251

QY 241 ICEDLLLP 249
Db ||||||||
252 ICEDLLLP 260

RESULT 3
AAU21747
ID AAU21747 standard; Protein; 260 AA.
XX
AC AAU21747;

XX 04-DEC-2001 (first entry)
XX Novel human neoplastic disease associated polypeptide #180.
DE
XX Human; neoplastic disease associated polypeptide; cancer;
KW hyperproliferative disorder; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW neuroprotective; cytostatic; anti inflammatory; vasotropic.
XX
OS Homo sapiens.
XX
PN WO200155163-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01358.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.

```
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465558/50.
DR N-PSDB; AAS34946.
XX
XX Novel polypeptides and polynucleotides useful as diagnostic reagents to
PT diagnose diseases or disorders associated with aberrant expression or
PT activity of polypeptides, and for treating cancers, rheumatoid
PT arthritis -
XX
XX Claim 11; SEQ ID No 474; 687pp; English.
XX
CC The present invention relates to the isolation of novel human neoplastic
CC disease associated polypeptides, and cDNA (AAS34767-AAS35050) and DNA
CC sequences encoding for these polypeptides. The sequences of the
CC invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of disorders involving neoplastic disease such as
CC hyperproliferative disorders (e.g. leukaemia, bone cancer, bladder
CC cancer, brain stem glioma, adult liver cancer, childhood cerebellar
CC astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may
CC also be useful for treating other disorders such as neural disorders,
CC immune system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC and renal disorders. The polynucleotide sequences of the invention are
CC also useful in gene therapy. AAU21568-AAU21851 represent the novel human
CC neoplastic disease associated polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 260 AA;

Query Match 100.0%; Score 1319; DB 22; Length 260;
Best Local Similarity 100.0%; Pred. No. 1.4e-121;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAVSLRLGDLVWGKLGYPWPVKIVNPPKDLKPRGKCKFFVKFFGTEDHAWIKVEQL 60
Db |||||
12 MAAVSLRLGDLVWGKLGYPWPVKIVNPPKDLKPRGKCKFFVKFFGTEDHAWIKVEQL 71
QY 61 KPYHAHKEEMIKNGKRFQQAQVDAVEEFLRRAGKDKQTSSHNSDDKNNRNSSEERSRP 120
Db |||||
72 KPYHAHKEEMIKNGKRFQQAQVDAVEEFLRRAGKDKQTSSHNSDDKNNRNSSEERSRP 131
QY 121 NSGDEKRKLSLSEGVKKNMGEKRVSSGSRGSKSPLKRAQEQSPRKGRPPPKDEKD 180
Db |||||
132 NSGDEKRKLSLSEGVKKNMGEKRVSSGSRGSKSPLKRAQEQSPRKGRPPPKDEKD 191
QY 181 LTIPESTVKGMMAGPMAAFKWOPTASEPVKDADPHFHFLLSQTEKPAVCYQAITKKLK 240
Db |||||
192 LTIPESTVKGMMAGPMAAFKWOPTASEPVKDADPHFHFLLSQTEKPAVCYQAITKKLK 251
QY 241 ICEDLLLP 249
Db |||||
252 ICEDLLLP 260.

RESULT 4
AAU21590
ID AAU21590 standard; Protein; 250 AA.
XX
```

```
AC AAU21590;
XX
DT 04-DEC-2001 (first entry)
XX
DE Novel human neoplastic disease associated polypeptide #23.
XX
KW Human; neoplastic disease associated polypeptide; cancer;
KW hyperproliferative disorder; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW neuroprotective; cytostatic; anti inflammatory; vasotropic.
XX
OS Homo sapiens.
XX
PN WO200155163-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01358.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
```


PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.

PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465558/50.
DR N-PSDB; AAS34789.
XX
PT Novel polypeptides and polynucleotides useful as diagnostic reagents to
PT diagnose diseases or disorders associated with aberrant expression or
PT activity of polypeptides, and for treating cancers, rheumatoid
PT arthritis
XX
PS Claim 11; SEQ ID No 317; 687pp; English.
XX
CC The present invention relates to the isolation of novel human neoplastic
CC disease associated polypeptides, and cDNA (AAS34767-AAS35050) and DNA
CC sequences encoding for these polypeptides. The sequences of the
CC invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of disorders involving neoplastic disease such as
CC hyperproliferative disorders (e.g. leukaemia, bone cancer, bladder
CC cancer, brain stem glioma, adult liver cancer, childhood cerebellar
CC astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may
CC also be useful for treating other disorders such as neural disorders,
CC immune system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC and renal disorders. The polynucleotide sequences of the invention are
CC also useful in gene therapy. AAU21568-AAU21851 represent the novel human
CC neoplastic disease associated polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 250 AA;

Query Match 99.6%; Score 1314; DB 22; Length 250;
Best Local Similarity 100.0%; Pred. No: 4e-121;
Matches 248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAVSLRLGDLVWGKLGRIYPPWPGKIVNPPKDLKKPRGKKCFVKKFFGTEDHAWIKVEQLK 61
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3 AAVSLRLGDLVWGKLGRIYPPWPGKIVNPPKDLKKPRGKKCFVKKFFGTEDHAWIKVEQLK 62
QY 62 PYHAHKEEMIKINKGRKFQQAQVDAVEEFLRRAGKGDQTSNHNSSDDKNRNSSEERSRN 121
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
63 PYHAHKEEMIKINKGRKFQQAQVDAVEEFLRRAGKGDQTSNHNSSDDKNRNSSEERSRN 122
QY 122 SGDEKRKLSLSEGVKKNMGEGKKRVSSGSSSERGSKSPLKRAQEQSPRKRGRPPKDEKDL 181
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
123 SGDEKRKLSLSEGVKKNMGEGKKRVSSGSSSERGSKSPLKRAQEQSPRKRGRPPKDEKDL 182
QY 182 TTPESSTVKGMMAGPMAAFKQPTASEPVKADDPHFHFFLLSQTKEPAVCYQAITKKLKI 241
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
183 TTPESSTVKGMMAGPMAAFKQPTASEPVKADDPHFHFFLLSQTKEPAVCYQAITKKLKI 242
QY 242 CEDLLLP 249
Db |||||||
243 CEDLLLP 250

RESULT 5
AAW69240
ID AAW69240 standard; Protein; 547 AA.

```
XX AAW69240;
AC
XX 21-OCT-1998 (first entry)
DT
XX Clone AQ73_3 protein sequence.
DE
XX
XX Secreted protein; nutritional source; cell proliferation activity;
KW cell differentiation activity; immune stimulant; tissue growth activator;
KW haematopoiesis regulator; anti-inflammatory; tumour invasion suppressor;
KW tumour inhibitor; clone AQ73_3.
XX
OS Homo sapiens.
XX WO9825962-A2.
PN
XX 18-JUN-1998.
PD
XX
XX 12-DEC-1997; 97WO-US23224.
PF
XX
XX 11-DEC-1997; 97US-0989232.
PR
XX 13-DEC-1996; 96US-0766263.
PR
XX
XX (GEMY ) GENETICS INST INC.
PA
XX
XX Agostino MJ, Jacobs K, Lavallie ER, Mccoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
PI
XX WPI; 1998-362424/31.
DR
DR N-PSDB; AAV44859.
DR
XX New isolated polynucleotides - obtained from human adult testis,
PT human adult ovary, human adult brain and human adult heart cDNA
PT libraries
PT
XX Claim 15; Page 66-67; 108pp; English.
PS
XX This sequence is secreted protein, encoded by a polynucleotide of the
CC invention. The DNA was isolated from a human adult ovary cDNA library,
CC and is designated clone AQ73_3. The DNA sequences and encoded
CC polypeptides can be used as nutritional sources or supplements, or may
CC exhibit e.g. cytokine and cell proliferation/differentiation activity,
CC immune stimulating or suppressing activity, haematopoiesis regulating
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC activin/inhibin activity, chemostatic/chemokinetic activity,
CC cadherin/tumour invasion suppressor activity, tissue growth activity,
CC tumour inhibition activity or other activities.
XX
XX Sequence 547 AA;
SQ
Query Match 97.9%; Score 1291; DB 19; Length 547;
Best Local Similarity 99.6%; Pred. No. 2.2e-118;
Matches 243; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAVSLRLGDLVWGKLGRIYPPWPGKIVNPPKDLKPRGKCKFFVKFFGTEHAWIKVEQL 60
DB 1 MAAVSLRLGDLVWGKLGRIYPPWPGKIVNPPKDLKPRGKCKFFVKFFGTEHAWIKVEQL 60
QY 61 KPYHAHKEEMIKKNGKRFQQAQVDAVEEFLRRAGKDKQTSSHNSDDKRRNSSEERSRP 120
DB 61 KPYHAHKEEMIKKNGKRFQQAQVDAVEEFLRRAGKDKQTSSHNSDDKRRNSSEERSRP 120
QY 121 NSGDEKRKLSLSEGVKKNMGEKKRVSSGSSERGSKSLKRAQEQSPRKRGRPPKDEK 180
DB 121 NSGDEKRKLSLSEGVKKNMGEKKRVSSGSSERGSKSLKRAQEQSPRKRGRPPKDEK 180
QY 181 LTIPESSTVKGMAGPMAAFKQWQPTASEPVKADPHFHFLLSQTEKPAVCYQAITKKLK 240
DB 181 LTIPESSTVKGMAGPMAAFKQWQPTASEPVKADPHFHFLLSQTEKPAVCYQAITKKLK 240
QY 241 ICED 244
DB 241 ICEE 244
```

```
RESULT 6
AAB90720
ID AAB90720 standard; Protein; 547 AA.
XX
AC AAB90720;
XX
DT 07-JUN-2001 (first entry)
XX
DE Human AQ73_3 protein sequence SEQ ID 137.
XX
KW Human; secreted protein; nutrient; cytokine modulator; proliferation;
KW differentiation; immune system modulator; tissue growth; chemotactic;
KW haemostatic; thrombolytic; anti-inflammatory; tumour inhibition;
KW haematopoiesis.
XX
OS Homo sapiens.
XX WO200119988-A1.
PN
XX 22-MAR-2001.
PD
XX 14-SEP-2000; 2000WO-US25135.
PF
XX 17-SEP-1999; 99US-0398829.
PR
XX (GEMY ) GENETICS INST INC.
PA
XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Bowman MR, Spaulding V, Agostino MJ;
PI
XX WPI; 2001-244801/25.
DR
DR N-PSDB; AAF98456.
DR
XX Isolated nucleic acids encoding polypeptides, useful for modulating
PT e.g. cytokine and cell proliferation/differentiation activity, the
PT immune system and hematopoiesis regulating activity -
XX
PS Disclosure; Page 468-470; 557pp; English.
XX
XX Human cDNA clones represented in AAF98374 - AAF98489 encode secreted
CC proteins AAB90667 - AAB90750. The cDNA clones are isolated from various
CC tissue types, and may be used in the prevention, treatment and diagnosis
CC of diseases associated with inappropriate protein expression. The
CC polypeptides and nucleic acids may be used as nutrients or to modulate
CC cytokine and cell proliferation/differentiation activity and may also be
CC involved in modulation of the immune system. The cDNA sequences,
CC proteins, their agonists and/or antagonists exhibit haematopoiesis
CC regulating activity; tissue growth activity; activin/inhibin activity;
CC chemotactic/chemokinetic activity; haemostatic and thrombolytic
CC activity; receptor/ligand activity; anti-inflammatory activity;
CC haematopoiesis activity; cadherin/tumour suppressor activity; and/or
CC tumour inhibition activity. Included in the invention are probes
CC represented in AAF98490 - AAF98572 which are specific for the cDNA clones
CC encoding the secreted proteins.
XX
SQ Sequence 547 AA;
Query Match 97.9%; Score 1291; DB 22; Length 547;
Best Local Similarity 99.6%; Pred. No. 2.2e-118;
Matches 243; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAVSLRLGDLVWGKLGRIYPPWPGKIVNPPKDLKPRGKCKFFVKFFGTEHAWIKVEQL 60
DB 1 MAAVSLRLGDLVWGKLGRIYPPWPGKIVNPPKDLKPRGKCKFFVKFFGTEHAWIKVEQL 60
QY 61 KPYHAHKEEMIKKNGKRFQQAQVDAVEEFLRRAGKDKQTSSHNSDDKRRNSSEERSRP 120
DB 61 KPYHAHKEEMIKKNGKRFQQAQVDAVEEFLRRAGKDKQTSSHNSDDKRRNSSEERSRP 120
QY 121 NSGDEKRKLSLSEGVKKNMGEKKRVSSGSSERGSKSLKRAQEQSPRKRGRPPKDEK 180
DB 121 NSGDEKRKLSLSEGVKKNMGEKKRVSSGSSERGSKSLKRAQEQSPRKRGRPPKDEK 180
```

Db 121 NSGDEKRLSLSEGKVKKNMGEGKRVSSGSSRGSKSPLKRAEQSPKRGPRPPKDEKD 180

Qy 181 LTIPESSTVKGMMAGPMAAFKQWPTASEPVKDADPHFHHLLSQTKEPAVCYQAITKKLK 240

Db 181 LTIPESSTVKGMMAGPMAAFKQWPTASEPVKDADPHFHHLLSQTKEPAVCYQAITKKLK 240

Qy 241 ICED 244

Db 241 ICEE 244

RESULT 7

AAO16415

ID AAO16415 standard; Protein; 553 AA.

AC AAO16415;

XX

DT 10-APR-2003 (first entry)

DE Human nucleic acid-associated protein (NAAP) - SEQ ID No 12.

XX

KW Human; nucleic acid-associated protein; NAAP; arteriosclerosis; cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS; cancer; developmental disorder; renal tubular acidosis; anaemia; asthma; mental retardation; neurological disorder; Alzheimer's disease; epilepsy; Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy; Crohn's disease; transgenic animal; animal model.

XX

OS Homo sapiens.

XX

PN WO2003000864-A2.

XX

PD 03-JAN-2003.

XX

PF 20-JUN-2002; 2002WO-US21179.

XX

PR 22-JUN-2001; 2001US-300518P.

PR 29-JUN-2001; 2001US-301787P.

PR 29-JUN-2001; 2001US-301792P.

PR 29-JUN-2001; 2001US-301892P.

PR 29-JUN-2001; 2001US-301893P.

PR 06-JUL-2001; 2001US-303405P.

PR 06-JUL-2001; 2001US-303442P.

PR 15-MAR-2002; 2002US-364438P.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Gandhi AR, Swarnakar A, Hafalia AJA, Warren BA, Emerling BM;

PI Arvizu CS, Ison CH, Honchell CD, Lee EA, Yue H, Forsythe IJ;

PI Ramkumar J, Griffin JA, Yang J, Sanjanwala MM, Baughn MR;

PI Borowsky ML, Yao MG, Walia NK, Bandman O, Lal PG, Becha SD;

PI Lee SY, Richardson TW, Elliott VS, Luo W, Tang YT, Zebbarjadian Y;

PI Lu Y;

XX

DR WPI; 2003-201420/19.

DR N-PSDB; AAL51565.

XX

PT New nucleic acid-associated proteins and polynucleotides, useful for diagnosing, treating or preventing cell proliferative (e.g. cancer), neurological (e.g. epilepsy or Parkinson's disease), or autoimmune disorders (e.g. AIDS)

XX

PS Claim 1; Page 227-228; 312pp; English.

XX

CC The invention comprises the amino acid and coding sequences of human nucleic acid-associated proteins (NAAP). The DNA and protein sequences of the invention are useful for diagnosing, treating or preventing disorders associated with aberrant expression of NAAP, such as: cell proliferative disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis or cancer); developmental disorders (e.g. renal tubular acidosis, anaemia or mental retardation); neurological disorders (e.g. Alzheimer's disease, Parkinson's disease or epilepsy); and autoimmune/inflammatory disorders (e.g. AIDS, allergies, asthma or Crohn's disease). The DNA sequences of

CC the invention are useful for creating transgenic animals to model human disease. The present amino acid sequence represents a human nucleic acid-associated protein of the invention.

SQ Sequence 553 AA;

Query Match 97.9%; Score 1291; DB 24; Length 553;
Best Local Similarity 99.6%; Pred. No. 2.2e-118;
Matches 243; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAAVSLRLGDLVWGKLGRLYPPWPVKIYNPPKDLKPRGKKCFVVKFPGTETHAWIKVEQL 60

Db 1 MAAVSLRLGDLVWGKLGRLYPPWPVKIYNPPKDLKPRGKKCFVVKFPGTETHAWIKVEQL 60

Qy 61 KPYHAHKEEMIKINKGRFQAVDAVEEFLRRRAKGDQTSNHNSSDDKNRRNSSEERSRP 120

Db 61 KPYHAHKEEMIKINKGRFQAVDAVEEFLRRRAKGDQTSNHNSSDDKNRRNSSEERSRP 120

Qy 121 NSGDEKRLSLSEGKVKKNMGEGKRVSSGSSRGSKSPLKRAEQSPKRGPRPPKDEKD 180

Db 121 NSGDEKRLSLSEGKVKKNMGEGKRVSSGSSRGSKSPLKRAEQSPKRGPRPPKDEKD 180

Qy 181 LTIPESSTVKGMMAGPMAAFKQWPTASEPVKDADPHFHHLLSQTKEPAVCYQAITKKLK 240

Db 181 LTIPESSTVKGMMAGPMAAFKQWPTASEPVKDADPHFHHLLSQTKEPAVCYQAITKKLK 240

Qy 241 ICED 244

Db 241 ICEE 244

RESULT 8

AAU21690

ID AAU21690 standard; Protein; 550 AA.

XX AC AAU21690;

XX DT 04-DEC-2001 (first entry)

DE Novel human neoplastic disease associated polypeptide #123.

XX Human; neoplastic disease associated polypeptide; cancer; hyperproliferative disorder; neural disorder; immune system disorder; muscular disorder; reproductive disorder; gastrointestinal disorder; pulmonary disorder; cardiovascular disorder; renal disorder; neuroprotective; cytostatic; anti inflammatory; vasotropic.

XX OS Homo sapiens.

XX PN WO200155163-A1.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01358.

XX PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465558/50.

N-PSDB; AAS34889.

Novel polypeptides and polynucleotides useful as diagnostic reagents to
diagnose diseases or disorders associated with aberrant expression or
activity of polypeptides, and for treating cancers, rheumatoid
arthritis -

Claim 11; SEQ ID No 417; 687pp; English.

The present invention relates to the isolation of novel human neoplastic
disease associated polypeptides, and cDNA (AAS34767-AAS35050) and DNA
sequences encoding for these polypeptides. The sequences of the
invention are useful in the diagnosis, treatment, prevention and/or
prognosis of disorders involving neoplastic disease such as
hyperproliferative disorders (e.g. leukaemia, bone cancer, bladder
cancer, brain stem glioma, adult liver cancer, childhood cerebellar
astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may
also be useful for treating other disorders such as neural disorders,
immune system disorders, muscular disorders, reproductive disorders,
gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
and renal disorders. The polynucleotide sequences of the invention are
also useful in gene therapy. AAU21568-AAU21851 represent the novel human

CC neoplastic disease associated polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 550 AA;
Query Match 96.9%; Score 1278; DB 22; Length 550;
Best Local Similarity 99.6%; Pred. No. 4.2e-117;
Matches 240; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 4 VSLRLGDLVWGKLGKRYPPWPGKIVNPPKDLKKPRGKCFVFKFGTETHAWIKVEQLKPY 63
Db 7 VSLRLGDLVWGKLGKRYPPWPGKIVNPPKDLKKPRGKCFVFKFGTETHAWIKVEQLKPY 66
Qy 64 HAHKEEMIKINKGRFQQAQVDAVEEFLRAKGDQTSNHNSSDDKNRRNSSEERSRPNNG 123
Db 67 HAHKEEMIKINKGRFQQAQVDAVEEFLRAKGDQTSNHNSSDDKNRRNSSEERSRPNNG 126
Qy 124 DEKRKLSLSEGVKKNMGEGKRVSSGSRGSKSPLKRAQEQSPRKRGPPPKDEKDLTI 183
Db 127 DEKRKLSLSEGVKKNMGEGKRVSSGSRGSKSPLKRAQEQSPRKRGPPPKDEKDLTI 186
Qy 184 PESSTVKGMMAGPMAAFKWOPTASEPVKDADPHFHHLLSQTEKPAVCYQAITKGLKICE 243
Db 187 PESSTVKGMMAGPMAAFKWOPTASEPVKDADPHFHHLLSQTEKPAVCYQAITKGLKICE 246
Qy 244 D 244
Db 247 E 247

RESULT 9
AAU20515
ID AAU20515 standard; Protein; 269 AA.
AC AAU20515;
XX
DT 06-DEC-2001 (first entry)
DE Human secreted protein, Seq ID No 507.
XX
KW Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
KW rheumatoid arthritis; antiarteriosclerotic; cardiac; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW Gaucher's disease; neurological disease; cerebrovascular disorder;
KW thrombosis; wound healing.
XX
OS Homo sapiens.
XX
PN WO200155326-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01347.
XX
PR 31-JAN-2000; 2000US-0179065.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-451931/48.
DR N-PSDB; AAS33224.
XX
PT New nucleic acids and polypeptides, useful for diagnosing, preventing
PT or treating medical conditions -
XX
PS Claim 11; SEQ ID No 507; 753pp; English.
XX
CC The invention relates to novel isolated nucleic acid molecules (I)

CC encoding human secreted proteins (II). (I) and (II) are used to prevent,
CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
CC the prevention, treatment and diagnosis of diseases associated with
CC inappropriate expression of secreted proteins. (I) and complementary
CC sequences may also be used as DNA probes in diagnostic assays (e.g.
CC polymerase chain reactions (PCR)) to detect and quantitate the presence
CC of similar nucleic acid sequences in samples, and so which patients may
CC be in need of restorative therapy. (II) may also be used as antigens in
CC the production of antibodies and in assays to identify modulators
CC (agonists and antagonists) of the expression and activity of the secreted
CC proteins. The anti-(II) antibodies and antagonists may also be used to
CC down regulate expression and activity of (II). The anti-(II) antibodies
CC may also be used as diagnostic agents for detecting the presence of (II)
CC in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
CC disorders include for example: immune/autoimmune diseases (e.g. HIV
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
CC angina and thrombosis), infections caused by bacteria, viruses and
CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),
CC agonists, antagonists and antibodies can also be used to promote wound
CC healing, maintain organs before transplantation, and support cell culture
CC of primary tissues. AAU20342-AAU20666 represent human secreted protein
CC amino acid sequences, and related sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification but was obtained in electronic format directly from WIPO
CC at: ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 269 AA;

Query Match 95.8%; Score 1263; DB 22; Length 269;
Best Local Similarity 98.8%; Pred. No. 4.7e-116;
Matches 238; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 4 VSLRLGDLVWGKLGKRYPPWPGKIVNPPKDLKKPRGKCFVFKFGTETHAWIKVEQLKPY 63
Db 1 VSLRLGDLVWGKLGKRYPPWPGKIVNPPKDLKKPRGKCFVFKFGTETHAWIKVEQLKPY 60
Qy 64 HAHKEEMIKINKGRFQQAQVDAVEEFLRAKGDQTSNHNSSDDKNRRNSSEERSRPNNG 123
Db 61 HAHKEEMIKINKGRFQQAQVDAVEEFLRAKGDQTSNHNSSDDKNRRNSSEERSRPNNG 120
Qy 124 DEKRKLSLSEGVKKNMGEGKRVSSGSRGSKSPLKRAQEQSPRKRGPPPKDEKDLTI 183
Db 121 DEKRKLSLSEGVKKNMGEGKRVSSGSRGSKSPLKRAQEQSPRKRGPPPKDEKDLTI 180
Qy 184 PESSTVKGMMAGPMAAFKWOPTASEPVKDADPHFHHLLSQTEKPAVCYQAITKGLKICE 243
Db 181 PESSTVKGMMAGPMAAFKWOPTASEPVKDADPHFHHLLSQTEKPAVCYQAITKGLKICE 240
Qy 244 D 244
Db 241 E 241

RESULT 10
AAU21813
ID AAU21813 standard; Protein; 269 AA.
XX
AC AAU21813;
XX
DT 06-DEC-2001 (first entry)
XX
DE Novel human neoplastic disease associated polypeptide #246.
XX
KW Human; neoplastic disease associated polypeptide; cancer;
KW hyperproliferative disorder; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;

XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-465558/50.
XX N-PSDB; AAS35012.
XX Novel polypeptides and polynucleotides useful as diagnostic reagents to
PT diagnose diseases or disorders associated with aberrant expression or
PT activity of polypeptides, and for treating cancers, rheumatoid
PT arthritis -
XX
XX Claim 11; SEQ ID No 540; 687pp; English.
XX
XX The present invention relates to the isolation of novel human neoplastic
CC disease associated polypeptides, and cDNA (AAS34767-AAS35050) and DNA
CC sequences encoding for these polypeptides. The sequences of the
CC invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of disorders involving neoplastic disease such as
CC hyperproliferative disorders (e.g. leukaemia, bone cancer, bladder
CC cancer, brain stem glioma, adult liver cancer, childhood cerebellar
CC astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may
CC also be useful for treating other disorders such as neural disorders,
CC immune system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC and renal disorders. The polynucleotide sequences of the invention are
CC also useful in gene therapy. AAU21568-AAU21851 represent the novel human
CC neoplastic disease associated polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 269 AA;

Query Match 95.8%; Score 1263; DB 22; Length 269;
Best Local Similarity 98.8%; Pred. No. 4.7e-116;
Matches 238; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 4 VSLRLGDLVWGKLGRIYPPWPGKIVNPPKDLKPRGKCFVFFGTEGTHAWIKVEQLKPY 63
DB 1 VSLXGLDLVWGKLGRIYPPWPGKIVNPPKDLKPRGKCFVFFGTEGTHAWIKVEQLKPY 60
QY 64 HAHKEEMIKINKGRFQQAQVDAVEEFLRRAGKDDQTSNHNSSDDKRRNSSSEERSRPSNG 123
DB 61 HAHKEEMIKINKGRFQQAQVDAVEEFLRRAGKDDQTSNHNSSDDKRRNSSSEERSRPSNG 120
QY 124 DEKRKLSLSEGVKKNMGEGKKRVSSGSSERSGSKSPLKRAQOSPRKRGRPPKDEKDLTI 183
DB 121 DEKRKLSLSEGVKKNMGEGKKRVSSGSSERSGSKSPLKRAQOSPRKRGRPPKDEKDLTI 180
QY 184 PESSTVKGMMAGPMAAFKQPTASEPVKDADPHFHLLSQTEKPAVCYQAITKKLICE 243
DB 181 PESSTVKGMMAGPMAAFKQPTASEPVKDADPHFHLLSQTEKPAVCYQAITKKLICE 240
QY 244 D 244
DB 241 E 241

RESULT 11
ABB58603
ID ABB58603 standard; Protein; 475 AA.
XX ABB58603;
AC ABB58603;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 2601.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

OS Drosophila melanogaster.
XX WO200171042-A2.
XX
XX 27-SEP-2001.
PD
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX N-PSDB; ABL02706.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 2601; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 475 AA;

Query Match 14.6%; Score 193; DB 22; Length 475;
Best Local Similarity 27.7%; Pred. No. 3.2e-10;
Matches 64; Conservative 32; Mismatches 89; Indels 46; Gaps 8;
QY 2 AAVSLRLGDLVWGKLGRIYPPWPGKIVNPPKDLKPRGKCFVFFGTEGTHAWIKVEQLK 61
DB 6 AAASYSIGDLVFAKVGKGYPPWPAKIT-----KSNNNKYNVYFYGTGTANIKLEDLF 58
QY 62 PYHAHKEEMI--KINKGRFQQAQVDAVEEFLR-----RAKGKDDQTSNHNSSDDKRR 111
DB 59 PYASNKERFATEKIMKRAKPIEAIDQIESALRGEDSAPIDLLDGAEPVAPPTTGD----- 113
QY 112 NSSEERSRPSNGDEKRLSLSEGVKKNMGEGKKRVSS-----GSSERSGSKSPLKRAQE 165
DB 114 GVKTEEPKPEGPPEPATAAPVAAAAAEKPKKSGTRKTKAPRHRVVDGSEAGAE----- 166
QY 166 QSPRKRGRPPKDEKDLTIPESSSTVKGMMAGPMAAFKQPTAS-EPVKDADP 215
DB 167 -----APPPAKRRVPTTEGLATAAAIPAAPAAAA---TPTSSKKS VKKSKP 207

RESULT 12
ABB72327
ID ABB72327 standard; Protein; 202 AA.
XX
XX ABB72327;
AC ABB72327;
XX
DT 04-APR-2002 (first entry)
XX
DE Rat protein isolated from skin cells SEQ ID NO: 651.
XX
KW Human; rat; mouse; skin cell; skin wound; cancer; growth defect;
KW developmental defect; inflammatory disease; dermatological; vulnerary;
KW immunomodulator; anti-inflammatory; cytostatic; neuroprotective.

appear to play a role in cell proliferative and developmental disorders.
A purified antagonist of HGFh can be administered to a subject as a
method of treating or preventing a cell proliferative disorder e.g.
atherosclerosis, hepatitis, and some cancers. Antibodies against HGFh and
antisense sequences may also be used as antagonist for treating the
above. Developmental disorders treated by HGFh include renal tubular
acidosis, anaemia, Cushing's syndrome, achondroplastic dwarfism,
Duchenne and Becker muscular dystrophy, and epilepsy. HGFh antibodies and
HGFh polynucleotides and polypeptides may also be used in various
diagnostic methods.

Sequence 203 AA;
Query Match 14.3%; Score 189; DB 20; Length 203;
Best Local Similarity 28.2%; Pred. No. 2.5e-10;
Matches 55; Conservative 37; Mismatches 79; Indels 24; Gaps 5;
QY 7 RLGLVWGKLGKRYPPWPGKIVNPPKDLKPRGKCCFFVVFEGTETHAWIKVEQLKPYHAH 66
Db 10 KAGDLVFAKMGYPHPWPARIDELPEGAVKPPANK-YPIFFGTHETAFGLPKDLFPYKEY 68
QY 67 KEEMIKINKGRFQQAQDAVE-----EFLRAKGDQTSNHNSSDDKRNNSSEERSRPN 121
Db 69 KDKFGKSNKRKGFNEGLWEIENNPVGKFTGYQAIQQSSSETEGEGGNTADASSE----- 123
QY 122 SGDEKRKLSLSEGVKKNMGEGKKRVSSGSSERGSPLKRAEQSPKRGPRPPKDEKDL 181
Db 124 -----EEGDRVEEDGKGRKNEKAGSKRKKSYSYTSKSKSQSRKSPG--DEDDKDC 171
QY 182 TTPES-STVKGMMAG 195
Db 172 KEENKSSSEGGDAG 186

RESULT 14
AAB13521
ID AAB13521 standard; Protein; 203 AA.
XX
AC AAB13521;
XX
DT 13-NOV-2000 (first entry)
XX
DE Human hepatoma-derived growth factor homologous polypeptide.
XX
KW Human; hepatoma-derived growth factor homologous polypeptide;
KW huHGFh; cell proliferation; haematopoiesis; angiogenesis;
KW lymphocyte proliferation; infection; autoimmune disease;
KW vascular disease; cancer.
XX
OS Homo sapiens.
XX
PN WO200037492-A2.
XX
PD 29-JUN-2000.
XX
PF 22-DEC-1999; 99WO-US30932.
XX
PR 22-DEC-1998; 98US-0113344.
XX
PA (ELIL) LILLY & CO ELI.
XX
PI Na S;
XX
DR WPI; 2000-442638/38.
DR N-PSDB; AAA65004.
XX
PT Novel human homologues of hepatoma-derived growth factor homologous
PT nucleic acids, polypeptides useful as probes or amplification primers
PT in the detection, quantitation or isolation of gene sequences or
PT transcripts -
XX
PS Claim 9; Page 74-75; 75pp; English.

The present sequence is the human homologue of the hepatoma-derived
growth factor (huHGFh). This protein is involved in tumour formation in
some cells, and it is thought that the gene and protein will be useful in
the diagnosis and treatment of infections, autoimmune disorders, vascular
diseases and cancers. In addition, the gene can be used to produce
transgenic animals which may be used as animal models for these diseases,
and the protein can be used to induce cell proliferation, haematopoiesis,
lymphocyte proliferation and angiogenesis.

Sequence 203 AA;
Query Match 14.3%; Score 189; DB 21; Length 203;
Best Local Similarity 28.2%; Pred. No. 2.5e-10;
Matches 55; Conservative 37; Mismatches 79; Indels 24; Gaps 5;
QY 7 RLGLVWGKLGKRYPPWPGKIVNPPKDLKPRGKCCFFVVFEGTETHAWIKVEQLKPYHAH 66
Db 10 KAGDLVFAKMGYPHPWPARIDELPEGAVKPPANK-YPIFFGTHETAFGLPKDLFPYKEY 68
QY 67 KEEMIKINKGRFQQAQDAVE-----EFLRAKGDQTSNHNSSDDKRNNSSEERSRPN 121
Db 69 KDKFGKSNKRKGFNEGLWEIENNPVGKFTGYQAIQQSSSETEGEGGNTADASSE----- 123
QY 122 SGDEKRKLSLSEGVKKNMGEGKKRVSSGSSERGSPLKRAEQSPKRGPRPPKDEKDL 181
Db 124 -----EEGDRVEEDGKGRKNEKAGSKRKKSYSYTSKSKSQSRKSPG--DEDDKDC 171
QY 182 TTPES-STVKGMMAG 195
Db 172 KEENKSSSEGGDAG 186

RESULT 15
AAY88374
ID AAY88374 standard; Protein; 203 AA.
XX
AC AAY88374;
XX
DT 20-JUL-2000 (first entry)
XX
DE Human type II hepatoma derived growth factor (HDGF2) amino acid sequence.
XX
KW Hepatoma-derived growth factor 2; HDGF2; cytostatic; drug;
KW treatment; hepatoma.
XX
OS Homo sapiens.
XX
PN WO200017351-A1.
XX
PD 30-MAR-2000.
XX
PF 06-SEP-1999; 99WO-CN00139.
XX
PR 22-SEP-1998; 98CN-0119758.
XX
PA (YULL/) YU L.
XX
PI Yu L, Zhang H, Fu Q, Zhao Y, Tu Q;
XX
DR WPI; 2000-283579/24.
DR N-PSDB; AAA13161.
XX
PT Type II human-derived growth factor with homology to its type I
PT version, useful e.g. in study and development of drugs for hepatoma -
XX
PS Claim 2; Page 15; 27pp; Chinese.
XX
CC This sequence represents the human hepatoma-derived growth factor (HDGF2)
CC protein sequence. The HDGF2 sequence shows considerable homology to the
CC type I version. The protein has cytostatic activity. The HDGF2 protein,
CC derivatives and polynucleotides are useful e.g. in the study and
CC development of drugs for the treatment of hepatoma.

```
SQ Sequence 203 AA;
Query Match 14.3%; Score 189; DB 21; Length 203;
Best Local Similarity 28.2%; Pred. No. 2.5e-10;
Matches 55; Conservative 37; Mismatches 79; Indels 24; Gaps 5;

QY 7 RLGDLVWGKLG RYPPWPGKIVNPPKDLKKPRGKKCFVVKFFGTEHAWIKVEQLKPYHAH 66
   : ||||: |: || || || |: || |: || |: || |: || |: || |: || |: || |:
Db 10 KAGDLVFAKMGYPHPWPARIDELPEGAVKPPANK-YPIFFFGTHETAFLGPKDLEPYKEY 68

QY 67 KEEMIKINKGRFQQAQDAVE-----EFLRRAKGDQDTSSHNSDDKNRRNSSEERSRPN 121
   |: || || || |: || |: || |: || |: || |: || |: || |: || |: || |:
Db 69 KDKFGKSNKRKGFNEGLWEIENNPVGKFTGYQAIQQQSSSETEGEGGNTADASSE----- 123

QY 122 SGDEKRKLSLSEGVKKNMGEKKRVSSGSSSERGSKSPLKRAQEQSPRKRGRPPKDEKDL 181
   || |: || |: || |: || |: || |: || |: || |: || |: || |: || |: ||
Db 124 -----EEGDRVEEDGKGRKNEKAGSKRKSYTSKSSKQSRKSPG--DEDDKDC 171

QY 182 TTPES-STVKGMMAG 195
   |: || |: ||
Db 172 KEENKSSSEGGDAG 186
```

Search completed: December 8, 2003, 11:09:13
Job time : 43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2003, 11:09:20 ; Search time 31 Seconds
(without alignments)
1493.869 Million cell updates/sec

Title: US-09-987-755-2
Perfect score: 1319
Sequence: 1 MAAVSLRLGDLVWGKLGKGRYP.....VCYQAITKKLKICEDLLLP 249

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1319	100.0	249	11	US-09-987-755-2
2	1319	100.0	260	15	US-10-103-313-474
3	1314	99.6	250	15	US-10-103-313-317
4	1291	97.9	547	12	US-10-067-482-3
5	1291	97.9	553	12	US-10-067-482-2
6	1278	96.9	550	15	US-10-103-313-417
7	1263	95.8	269	15	US-10-103-313-540
8	190.5	14.4	202	11	US-09-866-050A-651
9	189	14.3	203	9	US-09-768-826-41
10	189	14.3	203	12	US-10-247-671-132
11	186.5	14.1	246	9	US-09-925-302-511
12	185.5	14.1	221	12	US-10-418-445-4
13	185.5	14.1	237	9	US-09-938-885A-5
14	185.5	14.1	240	9	US-09-938-885A-3
15	185.5	14.1	240	11	US-09-987-755-7

16	185.5	14.1	240	12	US-10-116-275-220	Sequence 220, App
17	185.5	14.1	240	15	US-10-207-791-2	Sequence 2, Appli
18	178.5	13.5	235	9	US-09-938-885A-1	Sequence 1, Appli
19	165	12.5	314	9	US-09-768-826-43	Sequence 43, Appl
20	165	12.5	670	11	US-09-823-187-86	Sequence 86, Appl
21	165	12.5	670	12	US-09-863-776-51	Sequence 51, Appl
22	165	12.5	671	11	US-09-946-374-308	Sequence 308, App
23	165	12.5	671	11	US-09-823-187-85	Sequence 85, Appl
24	165	12.5	671	12	US-10-015-387A-308	Sequence 308, App
25	165	12.5	671	12	US-10-006-130A-308	Sequence 308, App
26	165	12.5	671	12	US-10-199-672-346	Sequence 346, App
27	165	12.5	671	12	US-10-006-172A-308	Sequence 308, App
28	165	12.5	671	12	US-10-187-749-346	Sequence 346, App
29	165	12.5	671	12	US-10-194-457-346	Sequence 346, App
30	165	12.5	671	12	US-10-184-642-346	Sequence 346, App
31	165	12.5	671	12	US-10-196-747-346	Sequence 346, App
32	165	12.5	671	12	US-10-015-392A-308	Sequence 308, App
33	165	12.5	671	12	US-10-017-253A-308	Sequence 308, App
34	165	12.5	671	12	US-10-173-689-346	Sequence 346, App
35	165	12.5	671	12	US-10-173-690-346	Sequence 346, App
36	165	12.5	671	12	US-10-173-691-346	Sequence 346, App
37	165	12.5	671	12	US-10-173-692-346	Sequence 346, App
38	165	12.5	671	12	US-10-173-694-346	Sequence 346, App
39	165	12.5	671	12	US-10-173-698-346	Sequence 346, App
40	165	12.5	671	12	US-10-173-699-346	Sequence 346, App
41	165	12.5	671	12	US-10-173-707-346	Sequence 346, App
42	165	12.5	671	12	US-10-174-569-346	Sequence 346, App
43	165	12.5	671	12	US-10-174-583-346	Sequence 346, App
44	165	12.5	671	12	US-10-174-587-346	Sequence 346, App
45	165	12.5	671	12	US-10-174-589-346	Sequence 346, App

ALIGNMENTS

RESULT 1

US-09-987-755-2
; Sequence 2, Application US/09987755
; Publication No. US20030022312A1
; GENERAL INFORMATION:
; APPLICANT: Kunsch et al
; TITLE OF INVENTION: Human Hepatoma-Derived Growth Factor-2
; FILE REFERENCE: PF198D1C1
; CURRENT APPLICATION NUMBER: US/09/987,755
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 09/263,625
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: 08/464,600
; PRIOR FILING DATE: 1995-06-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 249
; TYPE: PRT
; ORGANISM: human
US-09-987-755-2

Query Match	100.0%;	Score 1319;	DB 11;	Length 249;
Best Local Similarity	100.0%;	Pred. No. 1.3e-102;		
Matches	249;	Conservative	0;	Mismatches
		0;	Indels	0;
			Gaps	0;
Qy	1	MAAVSLRLGDLVWGKLGKGRYPWPWGKIVNPPKDLKKPRGKKCFVVKFFGTEDHAWIKVEQL	60	
Db	1	MAAVSLRLGDLVWGKLGKGRYPWPWGKIVNPPKDLKKPRGKKCFVVKFFGTEDHAWIKVEQL	60	
Qy	61	KPYHAHEMIKINKGRFQOAVDAVEEFLRRAKGDKDTSSHNSDDKNNRNSSEERSRP	120	
Db	61	KPYHAHEMIKINKGRFQOAVDAVEEFLRRAKGDKDTSSHNSDDKNNRNSSEERSRP	120	
Qy	121	NSGDEKRLSLSEGKVKKNMGEGKKRVSSGSSERGSPLKRAEQSPKRGRRPPKDEKD	180	
Db	121	NSGDEKRLSLSEGKVKKNMGEGKKRVSSGSSERGSPLKRAEQSPKRGRRPPKDEKD	180	

QY 181 LTIPESSTVKGMMAGPMAAFKQWQPTASEPVKADPHFHLLSQTTEKPAVCYQAITKKLK 240
Db 181 LTIPESSTVKGMMAGPMAAFKQWQPTASEPVKADPHFHLLSQTTEKPAVCYQAITKKLK 240
QY 241 ICEDLLLP 249
Db 241 ICEDLLLP 249
RESULT 2
US-10-103-313-474
; Sequence 474, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ207C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 474
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-313-474

Query Match 100.0%; Score 1319; DB 15; Length 260;
Best Local Similarity 100.0%; Pred. No. 1.4e-102;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAVSLRLGDLVWGKLGKRYPPWPGKIVNPPKDLKKPRGKCCFFVKGFGTETHAWIKVEQL 60
Db 12 MAAVSLRLGDLVWGKLGKRYPPWPGKIVNPPKDLKKPRGKCCFFVKGFGTETHAWIKVEQL 71
QY 61 KPYHAHKEEMIKKNGKRFQQAQVDAVEEFLRRAGKQDQTSNHNSSDDKNRRNSSEERSRP 120
Db 72 KPYHAHKEEMIKKNGKRFQQAQVDAVEEFLRRAGKQDQTSNHNSSDDKNRRNSSEERSRP 131
QY 121 NSGDEKRLSLSEGKVKKNMGEKGRVSSGSSERGSKSPKRAQEQSPRKRGRPPKDEK 180
Db 132 NSGDEKRLSLSEGKVKKNMGEKGRVSSGSSERGSKSPKRAQEQSPRKRGRPPKDEK 191
QY 181 LTIPESSTVKGMMAGPMAAFKQWQPTASEPVKADPHFHLLSQTTEKPAVCYQAITKKLK 240
Db 192 LTIPESSTVKGMMAGPMAAFKQWQPTASEPVKADPHFHLLSQTTEKPAVCYQAITKKLK 251
QY 241 ICEDLLLP 249
Db 252 ICEDLLLP 260

RESULT 3
US-10-103-313-317
; Sequence 317, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ207C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 317
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-313-317
Query Match 99.6%; Score 1314; DB 15; Length 250;

Best Local Similarity 100.0%; Pred. No. 3.4e-102;
Matches 248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 AAVSLRLGDLVWGKLGKRYPPWPGKIVNPPKDLKKPRGKCCFFVKGFGTETHAWIKVEQLK 61
Db 3 AAVSLRLGDLVWGKLGKRYPPWPGKIVNPPKDLKKPRGKCCFFVKGFGTETHAWIKVEQLK 62
QY 62 PYHAHKEEMIKKNGKRFQQAQVDAVEEFLRRAGKQDQTSNHNSSDDKNRRNSSEERSRP 121
Db 63 PYHAHKEEMIKKNGKRFQQAQVDAVEEFLRRAGKQDQTSNHNSSDDKNRRNSSEERSRP 122
QY 122 SGDEKRLSLSEGKVKKNMGEKGRVSSGSSERGSKSPKRAQEQSPRKRGRPPKDEK 181
Db 123 SGDEKRLSLSEGKVKKNMGEKGRVSSGSSERGSKSPKRAQEQSPRKRGRPPKDEK 182
QY 182 TPESTVKGMMAGPMAAFKQWQPTASEPVKADPHFHLLSQTTEKPAVCYQAITKKLK 241
Db 183 TPESTVKGMMAGPMAAFKQWQPTASEPVKADPHFHLLSQTTEKPAVCYQAITKKLK 242
QY 242 CEDLLLP 249
Db 243 CEDLLLP 250

RESULT 4
US-10-067-482-3
; Sequence 3, Application US/10067482
; Publication No. US20030148407A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies, Inc.
; TITLE OF INVENTION: Angiogenesis Dehydrogenase Gene
; FILE REFERENCE: 1U 102 R1
; CURRENT APPLICATION NUMBER: US/10/067,482
; CURRENT FILING DATE: 2002-02-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 547
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-067-482-3

Query Match 97.9%; Score 1291; DB 12; Length 547;
Best Local Similarity 99.6%; Pred. No. 7.5e-100;
Matches 243; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAVSLRLGDLVWGKLGKRYPPWPGKIVNPPKDLKKPRGKCCFFVKGFGTETHAWIKVEQL 60
Db 1 MAAVSLRLGDLVWGKLGKRYPPWPGKIVNPPKDLKKPRGKCCFFVKGFGTETHAWIKVEQL 60
QY 61 KPYHAHKEEMIKKNGKRFQQAQVDAVEEFLRRAGKQDQTSNHNSSDDKNRRNSSEERSRP 120
Db 61 KPYHAHKEEMIKKNGKRFQQAQVDAVEEFLRRAGKQDQTSNHNSSDDKNRRNSSEERSRP 120
QY 121 NSGDEKRLSLSEGKVKKNMGEKGRVSSGSSERGSKSPKRAQEQSPRKRGRPPKDEK 180
Db 121 NSGDEKRLSLSEGKVKKNMGEKGRVSSGSSERGSKSPKRAQEQSPRKRGRPPKDEK 180
QY 181 LTIPESSTVKGMMAGPMAAFKQWQPTASEPVKADPHFHLLSQTTEKPAVCYQAITKKLK 240
Db 181 LTIPESSTVKGMMAGPMAAFKQWQPTASEPVKADPHFHLLSQTTEKPAVCYQAITKKLK 240
QY 241 ICED 244
Db 241 ICEE 244

RESULT 5
US-10-067-482-2
; Sequence 2, Application US/10067482
; Publication No. US20030148407A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies, Inc.

;; TITLE OF INVENTION: Angiogenesis Dehydrogenase Gene
;; FILE REFERENCE: IU 102 R1
;; CURRENT APPLICATION NUMBER: US/10/067,482
;; CURRENT FILING DATE: 2002-02-07
;; NUMBER OF SEQ ID NOS: 4
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 2
;; LENGTH: 553
;; TYPE: PRT
;; ORGANISM: homo sapiens
US-10-067-482-2

Query Match 97.9%; Score 1291; DB 12; Length 553;
Best Local Similarity 99.6%; Pred. No. 7.6e-100;
Matches 243; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAAVSLRLGDLVWGKLGYPWPWGKI VNPDKLKKPRGKKCFVKKFFGTEHDHAWIKVEQL 60
Db |||||
Qy 1 MAAVSLRLGDLVWGKLGYPWPWGKI VNPDKLKKPRGKKCFVKKFFGTEHDHAWIKVEQL 60
Db |||||
Qy 61 KPYHAHKEEMIKINKGRFQQA VDAVEEFLRRAGKDKQTSSHNSSDDKNRRNSSEERSRP 120
Db |||||
Qy 61 KPYHAHKEEMIKINKGRFQQA VDAVEEFLRRAGKDKQTSSHNSSDDKNRRNSSEERSRP 120
Db |||||
Qy 121 NSGDEKRLSLSEGKVKKNMGEGKRVSSGSSERGSKSPDKRAEQSPRKRGRPPKDEKD 180
Db |||||
Qy 121 NSGDEKRLSLSEGKVKKNMGEGKRVSSGSSERGSKSPDKRAEQSPRKRGRPPKDEKD 180
Db |||||
Qy 181 LTIPESSTVKGMMAGPMAAFKQPTASEPVKDADPHFHHFLLSQTEKPAVCYQAITKKLK 240
Db |||||
Qy 181 LTIPESSTVKGMMAGPMAAFKQPTASEPVKDADPHFHHFLLSQTEKPAVCYQAITKKLK 240
Db |||||

Qy 241 ICED 244
Db |||| ICEE 244

RESULT 6
US-10-103-313-417
; Sequence 417, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJZ07C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 417
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-313-417

Query Match 96.9%; Score 1278; DB 15; Length 550;
Best Local Similarity 99.6%; Pred. No. 9.2e-99;
Matches 240; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 4 VSLRLGDLVWGKLGYPWPWGKI VNPDKLKKPRGKKCFVKKFFGTEHDHAWIKVEQLKPY 63
Db |||||
Qy 7 VSLRLGDLVWGKLGYPWPWGKI VNPDKLKKPRGKKCFVKKFFGTEHDHAWIKVEQLKPY 66
Db |||||
Qy 64 HAHKEEMIKINKGRFQQA VDAVEEFLRRAGKDKQTSSHNSSDDKNRRNSSEERSRPNSG 123
Db |||||
Qy 67 HAHKEEMIKINKGRFQQA VDAVEEFLRRAGKDKQTSSHNSSDDKNRRNSSEERSRPNSG 126
Db |||||
Qy 124 DEKRLSLSEGKVKKNMGEGKRVSSGSSERGSKSPDKRAEQSPRKRGRPPKDEKDLTI 183
Db |||||
Qy 127 DEKRLSLSEGKVKKNMGEGKRVSSGSSERGSKSPDKRAEQSPRKRGRPPKDEKDLTI 186
Db |||||
Qy 184 PESSTVKGMMAGPMAAFKQPTASEPVKDADPHFHHFLLSQTEKPAVCYQAITKKLKICE 243

Db 187 PESSTVKGMMAGPMAAFKQPTASEPVKDADPHFHHFLLSQTEKPAVCYQAITKKLKICE 246
Qy 244 D 244
Db 247 E 247

RESULT 7
US-10-103-313-540
; Sequence 540, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJZ07C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 540
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (165)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-103-313-540

Query Match 95.8%; Score 1263; DB 15; Length 269;
Best Local Similarity 98.8%; Pred. No. 6.9e-98;
Matches 238; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 VSLRLGDLVWGKLGYPWPWGKI VNPDKLKKPRGKKCFVKKFFGTEHDHAWIKVEQLKPY 63
Db |||||
Qy 1 VSLRLGDLVWGKLGYPWPWGKI VNPDKLKKPRGKKCFVKKFFGTEHDHAWIKVEQLKPY 60
Db |||||
Qy 64 HAHKEEMIKINKGRFQQA VDAVEEFLRRAGKDKQTSSHNSSDDKNRRNSSEERSRPNSG 123
Db |||||
Qy 61 HAHKEEMIKINKGRFQQA VDAVEEFLRRAGKDKQTSSHNSSDDKNRRNSSEERSRPNSG 120
Db |||||
Qy 124 DEKRLSLSEGKVKKNMGEGKRVSSGSSERGSKSPDKRAEQSPRKRGRPPKDEKDLTI 183
Db |||||
Qy 121 DEKRLSLSEGKVKKNMGEGKRVSSGSSERGSKSPDKRAEQSPRKRGRPPKDEKDLTI 180
Db |||||
Qy 184 PESSTVKGMMAGPMAAFKQPTASEPVKDADPHFHHFLLSQTEKPAVCYQAITKKLKICE 243
Db |||||
Qy 181 PESSTVKGMMAGPMAAFKQPTASEPVKDADPHFHHFLLSQTEKPAVCYQAITKKLKICE 240
Db |||||

RESULT 8
US-09-866-050A-651
; Sequence 651, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c4U


```
Query Match      14.1%; Score 186.5; DB 9; Length 246;
Best Local Similarity 29.8%; Pred. No. 7.8e-08;
Matches 64; Conservative 30; Mismatches 100; Indels 21; Gaps 8;

Qy 7 RLGDLVWGKLGKRYPPWPGKIVNPPKDLKPRGKCKFFVKFFGTHAWIKVEQLKPYHAH 66
Db 17 KCGDLVFAKMGYPHWPARDIEMPEAAVKSTANK-YQVFFFGTHETAFGLPKDLFPYEE 75

Qy 67 KEEMIKINKGRFQQAQVDAVEEFLRRRAKGDQTSNHNSSDDKNRRNSSEERSRPNNGDEK 126
Db 76 KEKFGKNKRKGFSEGLWEIEN-----NPTVKASGYQSSQKSCVVEPEPEAAEGDGD 130

Qy 127 RKLSEGVKKNMGEKRVSSGSSGSKSPKRAQ-----EQSPR--KRGRPPKDEKD 180
Db 131 KK-GNAEGSSDE---EGKLVIDEPAKEKNEKGALKRRAGDGLLEDSPKRPKEAENPEGEE- 185

Qy 181 LTIPESSTVKGMMAGPMAAFKQPTASEPVKDADP 215
Db 186 ---KEATLEVERPLPMEVEK-NSTXSEPGSGRGP 216

RESULT 12
US-10-418-445-4
; Sequence 4, Application US/10418445
; Publication No. US20030190708A1
; GENERAL INFORMATION:
; APPLICANT: BURGESS, et al.
; TITLE OF INVENTION: No. US20030190708A1el Hepatoma-Derived Growth Factor-Like Protein
; TITLE OF INVENTION: Polynucleotides Encoding Them and Methods of Use
; FILE REFERENCE: 15966-574 US
; CURRENT APPLICATION NUMBER: US/10/418,445
; CURRENT FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US/09/675,016
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: USSN 60/156975
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-418-445-4

Query Match      14.1%; Score 185.5; DB 12; Length 221;
Best Local Similarity 29.8%; Pred. No. 8.3e-08;
Matches 64; Conservative 30; Mismatches 100; Indels 21; Gaps 8;

Qy 7 RLGDLVWGKLGKRYPPWPGKIVNPPKDLKPRGKCKFFVKFFGTHAWIKVEQLKPYHAH 66
Db 2 KCGDLVFAKMGYPHWPARDIEMPEAAVKSTANK-YQVFFFGTHETAFGLPKDLFPYEE 60

Qy 67 KEEMIKINKGRFQQAQVDAVEEFLRRRAKGDQTSNHNSSDDKNRRNSSEERSRPNNGDEK 126
Db 61 KEKFGKNKRKGFSEGLWEIEN-----NPTVKASGYQSSQKSCVVEPEPEAAEGDGD 115

Qy 127 RKLSEGVKKNMGEKRVSSGSSGSKSPKRAQ-----EQSPR--KRGRPPKDEKD 180
Db 116 KK-GNAEGSSDE---EGKLVIDEPAKEKNEKGALKRRAGDGLLEDSPKRPKEAENPEGEE- 170

Qy 181 LTIPESSTVKGMMAGPMAAFKQPTASEPVKDADP 215
Db 171 ---KEATLEVERPLPMEVEK-NSTPSEPGSGRGP 201

RESULT 13
US-09-938-885A-5
; Sequence 5, Application US/09938885A
; Patent No. US20020090679A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Goli, Surya K.
```

```
Murty, Lynn E.
TITLE OF INVENTION: LUNG GROWTH FACTOR VARIANT
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/938,885A
FILING DATE: 24-Aug-2001
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/760,745
FILING DATE: 1996-12-07
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0169 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 237 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 945419
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-938-885A-5

Query Match      14.1%; Score 185.5; DB 9; Length 237;
Best Local Similarity 29.8%; Pred. No. 9.1e-08;
Matches 65; Conservative 32; Mismatches 94; Indels 27; Gaps 9;

Qy 7 RLGDLVWGKLGKRYPPWPGKIVNPPKDLKPRGKCKFFVKFFGTHAWIKVEQLKPYHAH 66
Db 11 KCGDLVFAKMGYPHWPARDIEMPEAAVKSTANK-YQVFFFGTHETAFGLPKDLFPYEE 69

Qy 67 KEEMIKINKGRFQQAQVDAVEEFLRRRAKGDQTSNHNSSDDKNRRNSSEERSRPNNGDEK 126
Db 70 KEKFGKNKRKGFSEGLWEIEN-----NPTVKASGYQSSQKSCAAEPEPEAHEGDGD 124

Qy 127 RKLSEGVKKNMGEKRVSSGSSGSKSPKRAQ-----AQEQSPKRGRRPK-----D 177
Db 125 KKGS-AEGSSDE---EGKLVIDEPAKEKNEKGTLKRRAGDVLDSPKR-----PKESGDHE 176

Qy 178 EKDLTIPESSTVKGMMAGPMAAFKQPTASEPVKDADP 215
Db 177 EED---KEIAALEGERHLPMEVEK-NSTPSEPDGSGQP 210

RESULT 14
US-09-938-885A-3
; Sequence 3, Application US/09938885A
; Patent No. US20020090679A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Goli, Surya K.
; Murty, Lynn E.
; TITLE OF INVENTION: LUNG GROWTH FACTOR VARIANT
; NUMBER OF SEQUENCES: 5
```

CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/938,885A
FILING DATE: 24-Aug-2001
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/760,745
FILING DATE: 1996-12-07

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0169 US

TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 240 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 598956

SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-938-885A-3

Query Match	14.1%;	Score 185.5;	DB 9;	Length 240;	
Best Local Similarity	29.8%;	Pred. No. 9.2e-08;			
Matches	64;	Conservative 30;	Mismatches 100;	Indels 21;	Gaps 8;
Qy	7	RLGDLVWGKLGRYPPWPGKIVNPPKDLKKPRGKKCFVKKFGTGDHAWIKVEQIKPYHAH	66		
Db	11	KCGDLVFAKMKGYPHWPARIDEMPEAAVKSTANK-YQVFFFGTHTAFLGPKDLFPYEE	69		
Qy	67	KEEMIKINKGRFQAVDAVEEFLRRRAKGDQTS SHNSSDDDKNRRNSSEERSRPSNGDEK	126		
Db	70	KEKFGKPNKRKGFSEGLWEIEN-----NPTVKASGYQSSQKKSCVEEPEPEAAEGDGD	124		
Qy	127	RKLSLSEGVKVMGEGKKRVSSGSSSERGSKSPLKRAQ-----EQSPR--KRGRPPKDEKD	180		
Db	125	KK-GNAEGSSDE---EGKLVIDEPAKEKNEKGALKRRAGDILLEDSPKRPKEAENPEGEE-	179		
Qy	181	LTIPESSTVKGMAGPMAAFKWQPTASEPVKDADP	215		
Db	180	---KEAATLEVERPLPMEVEK-NSTPSEPGSGRGP	210		

RESULT 15
US-09-987-755-7
; Sequence 7, Application US/09987755
; Publication No. US20030022312A1
; GENERAL INFORMATION:
; APPLICANT: Kunsch et al
; TITLE OF INVENTION: Human Hepatoma-Derived Growth Factor-2
; FILE REFERENCE: PF198D1C1
; CURRENT APPLICATION NUMBER: US/09/987,755
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 09/263,625
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: 08/464,600

```

; PRIOR FILING DATE: 1995-06-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 240
; TYPE: PRT
; ORGANISM: human
US-09-987-755-7

      Query Match      14.1%; Score 185.5; DB 11; Length 240;
      Best Local Similarity 29.8%; Pred. No. 9.2e-08;
      Matches 64; Conservative 30; Mismatches 100; Indels 21; Gaps 8;

Qy      7  RLGDLVWGKLG RYPPWPGKIVNPPKDLKPRGKKCFVVKFFGTEDHAWIKVEQLKPYHAH 66
Db      11  KCGDLVFAKMKGYPHWPARIDEMPEAAVKSTANK-YQVFFFGTHETAFLGPKDLFPYEES 69

Qy      67  KEEMIKINKGRFQQA VDAVEEFLRRAKGDQOTSSHNSDDKNRRNSSEERSRPNSGDEK 126
Db      70  KEKFGKPNKRKG FSEGLWEIEN-----NPTVKASGYQSSQKKSCVPEPEPEAAEGDGD 124

Qy      127  RKLSLSEGKVKGNMGEGKRVSSGSSERGSKSPLKRAQ-----EQSPR--KRGRPPKDEKD 180
Db      125  KK-GNAEGSSDE---EGKLVIDEPAKEKNEKGALKRRAGD LLEDSPKRPKEAENPEGEE- 179

Qy      181  LTIPESTVKGM MAGPMAAFKWQPTASEPVKDADP 215
Db      180  ---KEAATLEVERPLPMEVEK-NSTPSEPGSGRGP 210

```

Search completed: December 8, 2003, 11:12:21
Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2003, 11:07:05 ; Search time 35 Seconds
(without alignments)
1835.860 Million cell updates/sec

Title: US-09-987-755-2
Perfect score: 1319
Sequence: 1 MAAVSLRLGDLVWGKLGYP.....VCYQAITKKLKICEDLLLP 249

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertibrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1291	97.9	547	4 Q9BXK2	Q9bxk2 homo sapien
2	1247.5	94.6	546	11 Q922P9	Q922p9 mus musculus
3	1225.5	92.9	546	11 Q9CYQ1	Q9cyq1 mus musculus
4	1138	86.3	523	4 Q9BTI1	Q9bti1 homo sapien
5	201.5	15.3	239	6 Q9XSK7	Q9xsk7 bos taurus
6	193	14.6	475	5 Q9VAA9	Q9vaa9 drosophila
7	191.5	14.5	202	11 Q9JMG7	Q9jmg7 mus musculus
8	190.5	14.4	202	11 Q923W4	Q923w4 rattus norv
9	190.5	14.4	202	11 Q9D2M7	Q9d2m7 mus musculus
10	190.5	14.4	205	11 Q9BQ69	Q9bg69 mus musculus
11	189	14.3	203	4 Q9Y3E1	Q9y3e1 homo sapien
12	186.5	14.1	237	11 Q8BPG7	Q8bpg7 mus musculus
13	186.5	14.1	602	5 Q8T079	Q8t079 drosophila
14	186.5	14.1	603	5 Q9VL51	Q9vl51 drosophila
15	184.5	14.0	237	11 Q8VHK7	Q8vhk7 rattus norv
16	184.5	14.0	237	11 Q923W3	Q923w3 rattus norv

17	182.5	13.8	220	11 Q8BR62	Q8br62 mus musculu
18	165	12.5	670	4 Q9BW08	Q9bw08 homo sapien
19	164	12.4	283	13 Q8AVR7	Q8avr7 xenopus lae
20	160	12.1	331	11 Q99JF7	Q99jf7 mus musculu
21	160	12.1	528	11 Q99JF8	Q99jf8 mus musculu
22	157.5	11.9	304	3 Q06188	Q06188 saccharomyc
23	157.5	11.9	669	11 Q35540	Q35540 mus musculu
24	156.5	11.9	333	4 Q9NZI3	Q9nzi3 homo sapien
25	156.5	11.9	333	4 Q95368	Q95368 homo sapien
26	156.5	11.9	530	4 Q9UER6	Q9uer6 homo sapien
27	156.5	11.9	530	4 Q75475	Q75475 homo sapien
28	154	11.7	645	4 Q9H9W9	Q9h9w9 homo sapien
29	154	11.7	1388	4 Q9BYU8	Q9byu8 homo sapien
30	154	11.7	1437	4 Q9BYU9	Q9byu9 homo sapien
31	154	11.7	1437	4 Q9BZ95	Q9bz95 homo sapien
32	153.5	11.6	1449	10 Q8H3E3	Q8h3e3 oryza sativ
33	153	11.6	678	11 Q99L92	Q99l92 mus musculu
34	151.5	11.5	235	6 Q9XSK6	Q9xsk6 bos taurus
35	151.5	11.5	669	11 Q925G1	Q925g1 rattus norv
36	150	11.4	1392	10 Q9XER9	Q9xer9 arabidopsis
37	147	11.1	1445	10 Q9LEY4	Q9ley4 arabidopsis
38	142	10.8	1008	10 Q9FNE4	Q9fne4 arabidopsis
39	141.5	10.7	1358	11 Q8C2N9	Q8c2n9 mus musculu
40	140.5	10.7	1360	4 Q9BTB5	Q9btb5 homo sapien
41	140	10.6	387	11 Q8BJT3	Q8bjt3 mus musculu
42	137	10.4	832	5 Q95Q90	Q95q90 caenorhabdi
43	137	10.4	835	5 Q95Q91	Q95q91 caenorhabdi
44	136	10.3	726	10 Q9SF36	Q9sf36 arabidopsis
45	135.5	10.3	1016	5 Q8MT36	Q8mt36 drosophila

ALIGNMENTS

RESULT 1

Q9BXK2 ID Q9BXK2 PRELIMINARY; PRT; 547 AA.
AC Q9BXK2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Cytokine-like nuclear factor n-pac.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA New L., Han J.;
RT "A novel cytokine-like nuclear factor, n-pac."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF326966; AAK15524.1; .
DR InterPro; IPR006115; 6PGD_NAD.
DR InterPro; IPR000637; AT hook.
DR InterPro; IPR000313; PWWP_domain.
DR Pfam; PF02178; AT hook; 1.
DR Pfam; PF03446; NAD binding_2; 1.
DR SMART; SM00384; AT hook; 1.
DR PROSITE; PS50812; PWWP; 1.
SQ SEQUENCE 547 AA; 59827 MW; C7D785CCBF83204A CRC64;

Query Match 97.9%; Score 1291; DB 4; Length 547;
Best Local Similarity 99.6%; Pred. No. 1.5e-97;
Matches 243; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MAAVSLRLGDLVWGKLGYPWPWGKI VNPDKLKKPRGKCFVFKFFGTEDHAWIKVEQL 60
Db 1 MAAVSLRLGDLVWGKLGYPWPWGKI VNPDKLKKPRGKCFVFKFFGTEDHAWIKVEQL 60
Qy 61 KPYHAHKEIMIKNGKRFQQAQVDAVEEFLRAKGDQTS SHNSSDDDKNRNNSSEERSRP 120
Db 61 KPYHAHKEIMIKNGKRFQQAQVDAVEEFLRAKGDQTS SHNSSDDDKNRNNSSEERSRP 120


```
Q9D2M7
ID Q9D2M7 PRELIMINARY; PRT; 202 AA.
AC Q9D2M7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hepatoma-derived growth factor, related protein 3.
GN HDGFRP3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK019487; BAB31754.1; -
DR EMBL; AK044858; BAC32123.1; -
DR MGD; MGI:1352760; Hdgfrp3.
DR InterPro; IPR000313; PWWP_domain.
DR Pfam; PF00855; PWWP; 1.
DR SMART; SM00293; PWWP; 1.
DR PROSITE; PS50812; PWWP; 1.
SQ SEQUENCE 202 AA; 22430 MW; FCDB31BD2BEAB8D6 CRC64;

Query Match 14.4%; Score 190.5; DB 11; Length 202;
Best Local Similarity 28.7%; Pred. No. 5.7e-08;
Matches 56; Conservative 37; Mismatches 77; Indels 25; Gaps 6;

Qy 7 RLGLDVLWGKLGYPWPVKIYNPPKDLKPKRGKCFVFKFGTETHAWIKVEQLKPYHAH 66
: ||||: | : ||||: | : ||||: | : ||||: | : ||||: | : ||||: | : ||||:
Db 10 KAGDLVFAKMGYPHPARIDELPEGAVKPPANK-YPIFFFGTHTAFLGPKDLFPYKEY 68
|| : ||||: | : ||||: | : ||||: | : ||||: | : ||||: | : ||||:

Qy 67 KEEMIKINKGRFQQAQVADE-----EFLRAKGDQTSNHNSSDDKNRRNSSEERSRPN 121
|: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 69 KDKFGKSNKRKGFEGLWEIENPNPGVKFTGYQTIIQQSSSETEGEGGNTADASSE-----E 124
|| : ||||: | : ||||: | : ||||: | : ||||: | : ||||: | : ||||:

Qy 122 SGDEKRLKLSLSEGVKVKMNGEKKRVSSGSSRGSKSPLKRAEQSPRKGRPPKDEKDL 181
|| : ||||: | : ||||: | : ||||: | : ||||: | : ||||: | : ||||:
Db 125 EGD-----RVEDGKGRKNEKGGSKRKSYSYTSKSKQSRKSPG--DEDDKDC 170
|| : ||||: | : ||||: | : ||||: | : ||||: | : ||||: | : ||||:

Qy 182 TTPES-STVKGMMAG 195
|: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 171 KEENKSSSEGGDAG 185
|| : ||||: | : ||||: | : ||||: | : ||||: | : ||||: | : ||||:

RESULT 11
Q9Y3E1
ID Q9Y3E1 PRELIMINARY; PRT; 203 AA.
AC Q9Y3E1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CGI-142 protein (Hypothetical protein FLJ10418) (HRP-3) (Hepatoma-
DE derived growth factor 2).
GN HRP-3 OR HDGF2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Lin W.-C.;
RT "Comparative gene cloning: Identification of novel human genes with
RT Caenorhabditis elegans proteome as template.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
```

```
RESULT 10
Q8BQ69
ID Q8BQ69 PRELIMINARY; PRT; 205 AA.
AC Q8BQ69;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hepatoma-derived growth factor.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Dorsal root ganglion;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK051401; BAC34628.1; -
SQ SEQUENCE 205 AA; 22819 MW; E2ABAE3FBB31BD2B CRC64;

Query Match 14.4%; Score 190.5; DB 11; Length 205;
Best Local Similarity 28.7%; Pred. No. 5.8e-08;
Matches 56; Conservative 37; Mismatches 77; Indels 25; Gaps 6;

Qy 7 RLGLDVLWGKLGYPWPVKIYNPPKDLKPKRGKCFVFKFGTETHAWIKVEQLKPYHAH 66
: ||||: | : ||||: | : ||||: | : ||||: | : ||||: | : ||||: | : ||||:
Db 10 KAGDLVFAKMGYPHPARIDELPEGAVKPPANK-YPIFFFGTHTAFLGPKDLFPYKEY 68
|| : ||||: | : ||||: | : ||||: | : ||||: | : ||||: | : ||||:

Qy 67 KEEMIKINKGRFQQAQVADE-----EFLRAKGDQTSNHNSSDDKNRRNSSEERSRPN 121
|: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 69 KDKFGKSNKRKGFEGLWEIENPNPGVKFTGYQTIIQQSSSETEGEGGNTADASSE-----E 124
|| : ||||: | : ||||: | : ||||: | : ||||: | : ||||: | : ||||:

Qy 122 SGDEKRLKLSLSEGVKVKMNGEKKRVSSGSSRGSKSPLKRAEQSPRKGRPPKDEKDL 181
|| : ||||: | : ||||: | : ||||: | : ||||: | : ||||: | : ||||:
Db 125 EGD-----RVEDGKGRKNEKGGSKRKSYSYTSKSKQSRKSPG--DEDDKDC 170
|| : ||||: | : ||||: | : ||||: | : ||||: | : ||||: | : ||||:

Qy 182 TTPES-STVKGMMAG 195
|: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 171 KEENKSSSEGGDAG 185
|| : ||||: | : ||||: | : ||||: | : ||||: | : ||||: | : ||||:

RESULT 11
Q9Y3E1
ID Q9Y3E1 PRELIMINARY; PRT; 203 AA.
AC Q9Y3E1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CGI-142 protein (Hypothetical protein FLJ10418) (HRP-3) (Hepatoma-
DE derived growth factor 2).
GN HRP-3 OR HDGF2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Lin W.-C.;
RT "Comparative gene cloning: Identification of novel human genes with
RT Caenorhabditis elegans proteome as template.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
```



```

QY 126 KRKLSLSEGVKKNMGEGKRVSSGSSRG-----SKSPL-----KRAQEQS 167
Db 140 SEEIVGEEATADGNGVVAHV--GSPDEGDGLDVEINADSSASPVTSPTVTTKAAGKRT 197
QY 168 PR-----KRGRRPKDEKDLTIPESSTVKGMMAGPMAAFKQWPTAS-EPVKDAD 214
Db 198 PKAKSVAATSVKSTKGSAKSAQKRRTSAQQSP-----SGPSNAKRGKRDVSGEALQDAD 251

RESULT 14
Q9VL51 PRELIMINARY; PRT; 603 AA.
AC Q9VL51;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG4747 protein.
GN BEST:LD22483 OR CG4747.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA dePablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostali D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusekern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,

```

```

RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003627; AAF52846.2; -.
DR FlyBase; FBgn0043456; BEST:LD22483.
DR InterPro; IPR000313; PWWP_domain.
DR Pfam; PF00855; PWWP; 1.
DR SMART; SM00293; PWWP; 1.
DR PROSITE; PS50812; PWWP; 1.
SQ SEQUENCE 603 AA; 65354 MW; 98796823261F3D48 CRC64;

Query Match 14.1%; Score 186.5; DB 5; Length 603;
Best Local Similarity 25.5%; Pred.No. 4.3e-07;
Matches 61; Conservative 44; Mismatches 89; Indels 45; Gaps 10;

QY 10 DLVWGKLGYPWPWGKIVNPPKDL---KKPRGKKCFVKKFFGTEDHAWIKVEQLKPYHA- 65
Db 24 DLIWAKMGFTWPWGMIVDPDLDSQQRRANTKCVF--FFGSRNFAWIEENNIKPFEGP 81
QY 66 HKEEMIKNGKRFQQAQVDAVEEFLRRAKGDKQTSSHNSSDDKNRRNSSEERSRPNSGDE 125
Db 82 WKEELAKVSKPAAFRHAMTDIEKYIDDAEVDQVKNKSCGAPNHATEADFKIR--DGLD 139
QY 126 KRKLSLSEGVKKNMGEGKRVSSGSSRG-----SKSPL-----KRAQEQS 167
Db 140 SEEIVGEEATADGNGVVAHV--GSPDEGDGLDVEINADSSASPVTSPTVTTKAAGKRT 197
QY 168 PR-----KRGRRPKDEKDLTIPESSTVKGMMAGPMAAFKQWPTAS-EPVKDAD 214
Db 198 PKAKSVAATSVKSTKGSAKSAQKRRTSAQQSP-----SGPSNAKRGKRDVSGEALQDAD 251

RESULT 15
Q9VHK7 PRELIMINARY; PRT; 237 AA.
AC Q9VHK7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hepatoma-derived growth factor (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar;
RA Matsui H., Kawada N., Yokoya F., Takahara Y.;

```

RT "Isolation of rat HDGF from hepatic stellate cells."
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF448810; AAL47132.1; --
DR InterPro; IPR000313; PWWP_domain.

DR Pfam; PF00855; PWWP; 1.
DR SMART; SM00293; PWWP; 1.

DR PROSITE; PS50812; PWWP; 1.
FT NON TER 237 237

SQ SEQUENCE 237 AA; 26502 MW; C120F798BD40C1F0 CRC64;

Query Match 14.0%; Score 184.5; DB 11; Length 237;
Best Local Similarity 27.8%; Pred. No. 2.1e-07;
Matches 60; Conservative 35; Mismatches 98; Indels 23; Gaps 7;

QY	7	RLGDLVWGKLG	RYPPWPGKIVNPPKDLKKPRGKKCFVVKFFGTE	HAWIKVEQLKPYHAH	66
Db	11	KCGDLVFAKMGYP	HPARIDEMPEAAVKSTANK-YQVFFFGTHTAFLGPKDLFPYEE	S	69
QY	67	KEEMIKINKGRFQQA	VDAAVEEFLRRAKGDQTS	SHNSSDDKNRRNSSEERSRPN	SGDEK 126
Db	70	KEKFGKPNKRKG	FSGLWEIEN-----NPTVKASGYQSSQKSCAE	PEPEPEAHEGDGD	124
QY	127	RKLSLSEGVKVK	KMGEGKKRVSSSGSSRGSKSPLKRA-----	QEQSPRKRGPPPKDEK	179
Db	125	KK-GNAEGSSDE	---EGKLVIDEPAKEKNEKMGMLKRRAGDML	EDSPKRPKESGDHEEEK	180
QY	180	DLTIPESSTVKG	MMAGPMAAFKQWQPTASEPVKDADP	215	
Db	181	EI-----	AALEGERPLPVEMEK-NSTPSEPDSGGGP	210	

Search completed: December 8, 2003, 11:10:31
Job time : 37 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 8, 2003, 10:34:46 ; Search time 3727 Seconds
(without alignments)
2733.163 Million cell updates/sec

Title: US-09-987-755-2
Perfect score: 1319
Sequence: 1 MAAVSLRLGDLVWGKLGYP.....VCYQAITKKLKCEDLLLP 249

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09987755/runat_08122003_103435_18308/app_query.fasta_1.391
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09987755 @CGN 1 1 3508 @runat_08122003_103435_18308 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:
1: gb_ba:
2: gb_htg:
3: gb_in:
4: gb_om:
5: gb_ov:
6: gb_pat:
7: gb_ph:
8: gb_pl:
9: gb_pr:
10: gb_ro:
11: gb_sts:
12: gb_sy:
13: gb_un:
14: gb_vi:
15: em_ba:
16: em_fun:
17: em_hum:
18: em_in:
19: em_mu:
20: em_om:
21: em_or:
22: em_ov:
23: em_pat:
24: em_ph:
25: em_pl:
26: em_ro:
27: em_sts:
28: em_un:

29:	em_vi:*
30:	em_htg_hum:*
31:	em_htg_inv:*
32:	em_htg_other:*
33:	em_htg_mus:*
34:	em_htg_pln:*
35:	em_htg_rod:*
36:	em_htg_mam:*
37:	em_htg_vrt:*
38:	em_sy:*
39:	em_htgo_hum:*
40:	em_htgo_mus:*
41:	em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1291	97.9	1831	9	AF326966	AF326966 Homo sapi
2	1291	97.9	3316	6	AX099496	AX099496 Sequence
3	1291	97.9	3316	6	BD063979	BD063979 Secreted
4	1247.5	94.6	3187	10	BC006893	BC006893 Mus muscu
5	1138	86.3	2047	9	BC003693	BC003693 Homo sapi
6	915	69.4	140092	9	AL162615	AL162615 Human DNA
7	915	69.4	162847	2	AC025980	AC025980 Homo sapi
8	915	69.4	231260	2	AL160172	AL160172 Homo sapi
C 9	595.5	45.1	658	6	BD059639	BD059639 Secreted
C 10	551	41.8	154218	9	AC020663	AC020663 Homo sapi
C 11	551	41.8	169166	9	AC027687	AC027687 Homo sapi
C 12	514	39.0	233786	2	AC113755	AC113755 Rattus no
13	514	39.0	239081	2	AC108242	AC108242 Rattus no
14	514	39.0	243770	2	AC123492	AC123492 Rattus no
C 15	505.5	38.3	181475	10	AC127246	AC127246 Mus muscu
C 16	444	33.7	307	6	BD058512	BD058512 Secreted
17	341	25.9	445	6	BD045202	BD045202 Sequence
18	255.5	19.4	1893	3	AK114439	AK114439 Ciona int
19	204.5	15.5	4487	6	AR220855	AR220855 Sequence
20	201.5	15.3	1932	4	BTA237996	BTA237996 Bos tauru
21	198.5	15.0	69061	2	AC012986	AC012986 Drosophil
22	194.5	14.7	168469	3	AC007886	AC007886 Drosophil
23	194.5	14.7	181056	3	AC009733	AC009733 Drosophil
24	194.5	14.7	228443	3	AE003772	AE003772 Drosophil
25	193.5	14.7	1423	9	AF110642	AF110642 Homo sapi
26	193.5	14.7	1427	9	AB029156	AB029156 Homo sapi
27	193.5	14.7	1817	9	BC015483	BC015483 Homo sapi
28	193.5	14.7	1830	9	AF151900	AF151900 Homo sapi
29	193.5	14.7	1973	6	BD156109	BD156109 Primer fo
30	193.5	14.7	1973	9	AK001280	AK001280 Homo sapi
31	193.5	14.7	3805	6	BD131023	BD131023 Human gro
32	193	14.6	1859	3	AY051690	AY051690 Drosophil
33	191.5	14.5	612	10	AB029493	AB029493 Mus muscu
34	190.5	14.4	2750	10	AF389347	AF389347 Rattus no
35	190	14.4	2123	3	AY069497	AY069497 Drosophil
36	186.5	14.1	1744	10	BC005713	BC005713 Mus muscu
37	186.5	14.1	1752	10	AF251787	AF251787 Mus muscu
38	186.5	14.1	1763	10	BC021654	BC021654 Mus muscu
39	185.5	14.1	714	6	E13960	E13960 Mouse mRNA
40	185.5	14.1	723	6	E08545	E08545 DNA encodin
41	185.5	14.1	1563	6	E13961	E13961 Mouse mRNA
42	185.5	14.1	1563	10	MUSHDF	D63707 Mus musculu
43	185.5	14.1	2096	9	BC018991	BC018991 Homo sapi
44	185.5	14.1	2376	6	AR081931	AR081931 Sequence
45	185.5	14.1	2376	6	AX302495	AX302495 Sequence

ALIGNMENTS

```
AF326966
LOCUS AF326966 1831 bp mRNA linear PRI 04-MAR-2001
DEFINITION Homo sapiens cytokine-like nuclear factor n-pac mRNA, complete cds.
ACCESSION AF326966
VERSION AF326966.1 GI:13194723
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1831)
AUTHORS New,L. and Han,J.
TITLE A novel cytokine-like nuclear factor, n-pac
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1831)
AUTHORS New,L. and Han,J.
TITLE Direct Submission
JOURNAL Submitted (07-DEC-2000) Immunology, The Scripps Research Institute,
10550 North Torrey Pines Rd, La Jolla, CA 92037, USA
FEATURES
source
1..1831
/mol_type="mRNA"
/db_xref="taxon:9606"
1..1644
/codon_start=1
/product="cytokine-like nuclear factor n-pac"
/protein_id="AAK15524.1"
/db_xref="GI:13194724"
/translating="MAAVALRLGLVWKLGRYPWPVKIVNPPKDLKPRGKCPFFV
KFFGTEHAWIKVEQLPYHAHKEMIKNGKRFQQAQVDAVEFLRAKGDQTS
NSSDDKNRNSSESRPNSGDEKRLSLSEGKVKNGEGKRVSSGSSRSKSP
KRAEQSPKRGPRPPKDEKDLTIPESTVKGMMAGPMAAFKWPQTAPEPVKADPHF
HFLLSOTEKPAVCYQAITKKLKICEETGTSIQADSTAVNGSITPTDKKIGFLGL
LMGSGIVNLLKMGHTVWNRTAEKRGALRTPAEVSTCDITFACVSDPKAKDL
VLGPGSVLQGIKPGKCYDMSTVDADTVTELAQVIIVSRGGRFLEAPVSGNQQLSNDGM
LVILAAAGDRGLYEDCSFCQAMGKTSFELGEVGNAAKMLIVNMGVQSFMTIAEGLT
LAQVTGQSQQLLIDILNQGLASIFLDQKCNILQGNFKPDPFLYKIQKDLRLAIALG
DAVNHTPMAAAANEVYKRAKALDQSDNDMSAVRYAIIH"
BASE COUNT 487 a 472 c 510 g 362 t
ORIGIN
Alignment Scores:
Pred. No.: 7 26e-82 Length: 1831
Score: 1291.00 Matches: 243
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 97.88% Indels: 0
DB: 9 Gaps: 0
US-09-987-755-2 (1-249) x AF326966 (1-1831)
QY 1 MetAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
Db 1 ATGGCGGCTGTGAGTCTGCGGCTCGGCGACTTGGTGGGGGAACTCGGCCGATATCCT 60
QY 21 ProTrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLys 40
Db 61 CCTTGGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAACTCGCGGAAAGAAA 120
QY 41 CysPhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu 60
Db 121 TGCTTCTTTGTGAAATTTTGTGGAACAGAGATCATGCTGGATCAAAGTGGAAACAGCTG 180
QY 61 LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln 80
Db 181 AAGCCATATCATGCTCATAAAGAGGAAATGATAAAAAATTAAACAGGGTAAACGATTCAG 240
QY 81 GlnAlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSer 100
Db 241 CAAGCGGTAGATGTGTCGAAGATTCTCTCAGGAGAGCCAAAGGAAAGAACACGACGTCA 300
QY 101 SerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro 120
```

```
Db 301 TCCACAAATTTCTTCTGATGACAAAGAAATCGACGTAATTCAGTGAGGAGAGAGTAGGCCA 360
QY 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet 140
Db 361 AACTCAGGTGATGAGAAAGCGCAAACTTAGCCTGTCTGAAGGAAAGGTGAAGAAACATG 420
QY 141 GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu 160
Db 421 GGAGAAGGAAAGAGAGGGTGTCTTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG 480
QY 161 LysArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp 180
Db 481 AAAAGAGCCCAAGAGCAAAGTCCCGGAAGCGGGGTCCGCCCCCAAGAGGATGAGAAAG 540
QY 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
Db 541 CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCCGCTTT 600
QY 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
Db 601 AAATGGCAGCAACCGCAGCGAGCTGTAAAGATGCAGATCCTCATTTCCATCATTTTC 660
QY 221 LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys 240
Db 661 CTGCTAAGCAACACAGAGAAGCCAGCTGTCTGTACCAGCAATCACGAAGAAGTTGAAA 720
QY 241 IleCysGluAsp 244
Db 721 ATATGTGAAGAG 732
RESULT 2
AX099496 3316 bp DNA linear PAT 02-APR-2001
LOCUS AX099496
DEFINITION Sequence 136 from Patent WO0119988.
ACCESSION AX099496
VERSION AX099496.1 GI:13538581
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Jacobs,K., McCoy,J.M., Lavallie,E.R., Collins-Racie,L.A., Evans,C.,
Merberg,D., Treacy,M., Bowman,M.R., Spaulding,V. and Agostino,M.J.
TITLE Secreted proteins and polynucleotides encoding them
JOURNAL Patent: WO 0119988-A 136 22-MAR-2001;
Genetics Institute, Inc. (US)
FEATURES
Location/Qualifiers
1..3316
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 875 a 813 c 887 g 741 t
ORIGIN
Alignment Scores:
Pred. No.: 1 39e-81 Length: 3316
Score: 1291.00 Matches: 243
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 97.88% Indels: 0
DB: 6 Gaps: 0
US-09-987-755-2 (1-249) x AX099496 (1-3316)
QY 1 MetAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
Db 7 ATGGCGGCTGTGAGTCTGCGGCTCGGCGACTTGGTGGGGGAACTCGGCCGATATCCT 66
QY 21 ProTrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLys 40
Db 67 CCTTGGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAA 126
```

Qy	41	CysPhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu	60
Db	127	TGCTTCTTTGTGAAATTTTTTGGAAACAGAAAGATCATGCTGGATCAAAGTGGAACAGCTG	186
Qy	61	LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln	80
Db	187	AAGCCATATCATGCTCATAAAGAGGAAATGATAAAAATTAAACAAGGTAAACGATTCCAG	246
Qy	81	GlnAlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSer	100
Db	247	CAAGCGGTAGATGCTGTCTGAAGAGTCTCTCAGGAGAGCCAAAGGGAAAGACCAGACGTCA	306
Qy	101	SexHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro	120
Db	307	TCCACAAATCTTCTGATGACAAAGATCGACGTAATTCAGTGAGGAGAGAAAGTAGGCCA	366
Qy	121	AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet	140
Db	367	AACTCAGGTGATGAGAAAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAGACATG	426
Qy	141	GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu	160
Db	427	GGAGAAAGGAAAGAGAGGGGTGCTTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG	486
Qy	161	LysArgAlaGlnGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp	180
Db	487	AAAAGAGCCCAAGAGCAAAGTCCCCGGAAGCGGGGTGCGCCCCCAAGGATGAGAAGGAT	546
Qy	181	LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe	200
Db	547	CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCGTTT	606
Qy	201	LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisHisPhe	220
Db	607	AAATGGCAGCCAAACCGCAAGCGAGCGCTGTTAAAGATGCAGATCCTCATTTCCATCATTTT	666
Qy	221	LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys	240
Db	667	CTGTAAAGCCAAACAGAGAAGCCAGCTGTCTGTTACAGGCAATCACGAAGAAGTTGAAA	726
Qy	241	IleCysGluAsp	244
Db	727	ATATGTGAAGAG	738

RESULT 3	
BD063979	
LOCUS	3316 bp DNA linear PAT 27-AUG-2002
DEFINITION	Secreted proteins and polynucleotides encoding them.
ACCESSION	BD063979
VERSION	BD063979.1 GI:22609582
KEYWORDS	JP 2001506848-A/3.
SOURCE	unidentified
ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 3316)
AUTHORS	Jacobs, K., Mccoy, J.M., Lavallie, E.R., Racie, L.A., Merberg, D., Trecy, M., Spaulding, V. and Agostino, M.J.
TITLE	Secreted proteins and polynucleotides encoding them
JOURNAL	Patent: JP 2001506848-A 3 29-MAY-2001;
COMMENT	GENETICS INSTITUTE INC PN JP 2001506848-A/3 PD 29-MAY-2001 PF 12-DEC-1997 JP 1998525996 PR 13-DEC-1996 US 08/766263 PI KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A RACIE, PI DAVID MERBERG, PI MAURICE TREACY, VIKKI SPAULDING, MICHAEL J AGOSTINO PC C12N15/12, C12N5/10, C07K14/47, A61K38/17 CC Strandedness: Double; CC Topology: Linear; FH Key Location/Qualifiers.
FEATURES	Location/Qualifiers

repeat_region	MLT1G repeat: matches 127. .512 of consensus"			
repeat_region	14887. .15197			
repeat_region	/note="AluSx repeat: matches 1. .302 of consensus"			
repeat_region	15314. .15370			
repeat_region	/note="L1ME3 repeat: matches 5527. .5583 of consensus"			
repeat_region	L1MC/D repeat: matches 5527. .5583 of consensus"			
repeat_region	15549. .15874			
repeat_region	/note="7SLRNA repeat: matches 1. .319 of consensus"			
repeat_region	7SLRNA repeat: matches 1. .319 of consensus"			
repeat_region	15884. .16168			
repeat_region	/note="AluSx repeat: matches 3. .313 of consensus"			
repeat_region	16317. .16632			
repeat_region	/note="AluSx repeat: matches 1. .297 of consensus"			
repeat_region	complement(16709. .17162)			
repeat_region	/note="match: GSS: Em:AQ628320"			
repeat_region	17164. .17592			
repeat_region	/note="match: GSS: Em:AQ469272"			
repeat_region	17231. .17359			
repeat_region	/note="L1ME3 repeat: matches 6042. .6164 of consensus"			
repeat_region	L1ME3 repeat: matches 6042. .6164 of consensus"			
repeat_region	17435. .17528			
repeat_region	/note="MER47A repeat: matches 4. .99 of consensus"			
repeat_region	MER47A repeat: matches 4. .99 of consensus"			
repeat_region	17529. .17835			
repeat_region	/note="AluSx repeat: matches 1. .307 of consensus"			
repeat_region	17836. .18113			
repeat_region	/note="MER47A repeat: matches 99. .366 of consensus"			
repeat_region	MER47A repeat: matches 99. .366 of consensus"			
repeat_region	18116. .18317			
repeat_region	/note="L1ME3 repeat: matches 5697. .5898 of consensus"			
repeat_region	L1ME3 repeat: matches 5697. .5898 of consensus"			
repeat_region	18981. .19288			
repeat_region	/note="AluSx repeat: matches 1. .305 of consensus"			
repeat_region	19723. .19853			
repeat_region	/note="FLAM_C repeat: matches 19. .121 of consensus"			
repeat_region	19876. .20178			
repeat_region	/note="AluSx repeat: matches 1. .304 of consensus"			
repeat_region	20203. .20241			
repeat_region	/note="MIR repeat: matches 203. .241 of consensus"			
repeat_region	20255. .20553			
repeat_region	/note="AluSx repeat: matches 1. .299 of consensus"			
Alignment Scores:				
Pred. No.:	2.15e-53	Length:	140092	
Score:	915.00	Matches:	204	
Percent Similarity:	63.04%	Conservative:	16	
Best Local Similarity:	58.45%	Mismatches:	24	
Query Match:	69.37%	Indels:	108	
DB:	9	Gaps:	.1	
US-09-987-755-2 (1-249) x AL162615 (1-140092)				
Qy	1	MetAlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro	20	
Db	122725	ATGGGGCTGTGAGTTCGCTGCTCAGCAACTGTGATGGGGGAAACTCGGCTGTATCCT	122784	
Qy	21	ProTrpProGlyLysIleValAsnProProLysAspLeuLysLysProArg	37	
Db	122785	CCTTACCAGAAAGATCGTTAATCCACCGAAGGACTTGAGAAACCTCA-TGGGGCTGG	122843	
Qy	37	-----	37	
Db	122844	GCGCGTGGTCAAGCCTGTGATCACAGCACTTTGGGAGGCCAAGGTGGCGGATCACGA	122903	
Qy	37	-----	37	
Db	122904	GGTCAGGAGATTGAGACCATCCTGGCTAACATGGTGAAATCCCGTCTATATACTAAAAATAC	122963	
Qy	37	-----	37	
Db	122964	AAAAAAAATAATAGCCAGGCGTGGTGGCGGCGTCTAGTCTAGCTCCTCAGGAGGCT	123023	
Qy	37	-----	37	

Db	123024	GAGGCAGAGAAATAGCATGAACCCGGGAGGTGGAGCTTGCACTGAGCCGAGATCGCACCA	123083
Qy	37	-----	37
Db	123084	CTGCACCTCAACCTGGGCAACAGAGTGAGACTCCATCTCAAAAAAAGAAACCTCAT	123143
Qy	38	GlyLysLysCysPhePheValLys-PhePheGlyThrGluAspHisAlaTrpIleLysVa	57
Db	123144	GGAAAGAAATGCTTCTTGTGAAATTTTGTGAACATAAGATCATGCTGATTAAGT	123203
Qy	57	lGluGlnLeuLysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLy	77
Db	123204	GAAACAGCTGAAGCCATATCATGCTCATAAAGAGGAATGATAAGATTACAAGGGTAA	123263
Qy	77	sArgPheGlnGlnAlaValAspAlaValGluGluPheLeuArgAla-LysGlyLysA	97
Db	123264	ACTATTCAGCAAGCGGTGGATGCCGCCAAAGAGTTTCTCAGGAGAGCCCAAGGAAAG	123323
Qy	97	spGlnThrSerSerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluA	117
Db	123324	ATCAGACGTCAATCCCAATTTCTGCTGATGACAAGAATCGGCTTAATCCAGTGAGGAGA	123383
Qy	117	rgSerArgProAsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValL	137
Db	123384	GAAGCAGGCCAAACTCAGGTGATGAGAAGTGCAAAAC--AGCCTGTCTGAAGGAAAGTGA	123441
Qy	137	ysLysAsnMetGlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerL	157
Db	123442	AGAAGATCATGGGAGATGGAAAGAGGGGGTGTCTTCGGGCTCTTCAGAGACAGGCTCCA	123501
Qy	157	ysSerProLeuLysArgAlaGlnGlnSerProArgLysArgGlyArgProProLysA	177
Db	123502	AATCCCTCTGAAAAGAATCCAAAGAGCAAAAGTCCCCAGAGCGGGGTGACGCCCCCAAGA	123561
Qy	177	spGluLysAspLeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProM	197
Db	123562	ATGAGAGAAATGTCAACATCGTGGAGTCGGGTACCTTGAAGTGGATGATGCGCGGACCCAG	123621
Qy	197	etAlaAlaPheLysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisP	217
Db	123622	TGGCTGCATTTAAATGGCAGCCGACCAAGAGCGAGCGCTGTAAAGATGCAGATCCTCAT	123681
Qy	217	heHisHisPheLeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrL	237
Db	123682	TCCATCATTTCTGCTAAGCCAAACCAAGCCATC--TCTGTCTGTTACCAGGCAATCACAA	123739
Qy	237	ysLysLeuLysIleCysGluAsp	244
Db	123740	AGAAATTGAAAATATGTGAAGAG	123762
RESULT 7			
AC025980			
LOCUS			
DEFINITION			
AC025980			
AC025980.3			
HTG; HTGS PHASE1; HTGS_DRAFT.			
Homo sapiens (human)			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			

AC025980 162847 bp DNA linear HTG 07-JUN-2000
Homo sapiens chromosome 20 clone RP11-444K9 map 20, WORKING DRAFT
SEQUENCE, 30 unordered pieces.
AC025980
AC025980.3 GI:8317124
HTG; HTGS PHASE1; HTGS_DRAFT.
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 162847)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,

Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (18-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 7, 2000 this sequence version replaced gi:7652052.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6971
Center clone name: 444_K_9
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 150127 bases at least Q40
Consensus quality: 156059 bases at least Q30
Consensus quality: 158474 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 159947; sum-of-contigs
Quality coverage: 4.2 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1179: contig of 1179 bp in length
1180 1279: gap of 100 bp
1280 2395: contig of 1116 bp in length
2396 2495: gap of 100 bp
2496 3581: contig of 1086 bp in length
3582 3681: gap of 100 bp
3682 4980: contig of 1299 bp in length
4981 5080: gap of 100 bp
5081 6590: contig of 1510 bp in length
6591 6690: gap of 100 bp
6691 8194: contig of 1504 bp in length
8195 8294: gap of 100 bp
8295 9957: contig of 1663 bp in length
9958 10057: gap of 100 bp
10058 11940: contig of 1883 bp in length
11941 12040: gap of 100 bp
12041 13442: contig of 1402 bp in length
13443 13542: gap of 100 bp
13543 16309: contig of 2767 bp in length
16310 16409: gap of 100 bp
16410 18496: contig of 2087 bp in length
18497 18596: gap of 100 bp

18597 21162: contig of 2566 bp in length
21163 21262: gap of 100 bp
21263 24271: contig of 3009 bp in length
24272 24371: gap of 100 bp
24372 26834: contig of 2463 bp in length
26835 26934: gap of 100 bp
26935 29054: contig of 2120 bp in length
29055 29154: gap of 100 bp
29155 32187: contig of 3033 bp in length
32188 32287: gap of 100 bp
32288 35290: contig of 3003 bp in length
35291 37991: contig of 2601 bp in length
37992 38091: gap of 100 bp
38092 41561: contig of 3470 bp in length
41562 41661: gap of 100 bp
41662 48058: contig of 6397 bp in length
48059 48158: gap of 100 bp
48159 52403: contig of 4245 bp in length
52404 52503: gap of 100 bp
52504 58539: contig of 6036 bp in length
58540 58639: gap of 100 bp
58640 64768: contig of 6129 bp in length
64769 64868: gap of 100 bp
64869 74871: contig of 10003 bp in length
74872 74971: gap of 100 bp
74972 85271: contig of 10300 bp in length
85272 85371: gap of 100 bp
85372 94554: contig of 9183 bp in length
94555 94654: gap of 100 bp
94655 102707: contig of 8053 bp in length
102708 102807: gap of 100 bp
102808 116262: contig of 13455 bp in length
116263 116362: gap of 100 bp
116363 134631: contig of 18269 bp in length
134632 134731: gap of 100 bp
134732 162847: contig of 28116 bp in length.

FEATURES

Location/Qualifiers
1. 162847
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="20"
/map="20"
/clone="RP11-444K9"
/clone_lib="RPCI-11 Human Male BAC"
1. 1179
/note="assembly_fragment"
misc_feature 1280. 2395
/note="assembly_fragment"
misc_feature 2496. 3581
/note="assembly_fragment"
misc_feature 3682. 4980
/note="assembly_fragment"
misc_feature 5081. 6590
/note="assembly_fragment"
misc_feature 6691. 8194
/note="assembly_fragment"
misc_feature 8295. 9957
/note="assembly_fragment"
misc_feature 10058. 11940
/note="assembly_fragment"
misc_feature 12041. 13442
/note="assembly_fragment"
misc_feature 13543. 16309
/note="assembly_fragment"
misc_feature 16410. 18496
/note="assembly_fragment"
misc_feature 18597. 21162
/note="assembly_fragment"
misc_feature 21263. 24271
/note="assembly_fragment"
misc_feature 24372. 26834

misc_feature /note="assembly_fragment"
26935..29054
/note="assembly_fragment
clone_end:T7
vector_side:right"
misc_feature 29155..32187
/note="assembly_fragment"
misc_feature 32288..35290
/note="assembly_fragment"
misc_feature 35391..37991
/note="assembly_fragment"
misc_feature 38092..41561
/note="assembly_fragment"
misc_feature 41662..48058
/note="assembly_fragment"
misc_feature 48159..52403
/note="assembly_fragment"
misc_feature 52504..58539
/note="assembly_fragment"
misc_feature 58640..64768
/note="assembly_fragment"
misc_feature 64869..74871
/note="assembly_fragment"
misc_feature 74972..85271
/note="assembly_fragment"
misc_feature 85372..94554
/note="assembly_fragment"
misc_feature 94655..102707
/note="assembly_fragment
clone_end:SP6
vector_side:right"
misc_feature 102808..116262
/note="assembly_fragment"
misc_feature 116363..134631

Alignment Scores:
Pred. No.: 2.53e-53 Length: 162847
Score: 915.00 Matches: 204
Percent Similarity: 63.04% Conservative: 16
Best Local Similarity: 58.45% Mismatches: 24
Query Match: 69.37% Indels: 108
DB: 2 Gaps: 1

US-09-987-755-2 (1-249) x AC025980 (1-162847)

Qy 1 MetAlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgFyrPro 20
Db 24776 ATGGGGCTGTGAGTTCGCTGCCTCAGCAACTGTATGTGGGGAACTCGGTATCCT 24835
Qy 21 ProTrpProGlyLysIleValAsnProProLysAspLeuLysLysProArg----- 37
Db 24836 CCTTCACCAGGAAGATCGTTAATCCACCGAAGGACTTGAAGAAACCTCA-TGGGGCTGG 24894
Qy 37 ----- 37
Db 24895 GCGCGTGGCTCAGCCCTGTGATCACAGCACTTTGGGAGGCCAAGGTGGCGGATCACGA 24954
Qy 37 ----- 37
Db 24955 GGTCAGGAGATTGAGACCATTCTGGCTAACATGGTGAATCCCGTCTATATACTAAAAATAC 25014
Qy 37 ----- 37
Db 25015 AAAAAAATAATTAGCCAGCGGTGGTGGCGGCGTCTAGTCTCTCAGGAGGCT 25074
Qy 37 ----- 37
Db 25075 GAGGAGGAGATAGCATGAACCCGGGAGGTGGAGCTTGAGTGAGCCGAGATCGCACCA 25134
Qy 37 ----- 37
Db 25135 CTGCACTCAACCTGGGCAACAGAGTGAGACTCCATCTCAAAAAAAGAACCTCAT 25194

Qy 38 GlyLysLysCysPhePheValLys-PhePheGlyThrGluAspHisAlaTrpIleLysVa 57
Db 25195 GGAAGAAATGCTTCTTTGTGAATTTTGTGAACATAAGATCATGCTGGATTAAAGT 25254
Qy 57 lGluGlnLeuLysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLy 77
Db 25255 GAAACAGCTGAAGCATATCATGTCTATAAAGAGAAATGATAAAGATTCAAGGGTAA 25314
Qy 77 sArgPheGlnGlnAlaValAspAlaValGluGluPheLeuArgArgAla-LysGlyLysA 97
Db 25315 ACTATTCCAGCAAGCGTGGATGCCGCCAAAGAGTTTCTCAGGAGAGCCCAAGGGAAG 25374
Qy 97 spGlnThrSerSerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluA 117
Db 25375 ATCAGACGTATCCCAATTTCTGCTGATGACAAGAATCGGCTTAATTCCAGTGAGAGA 25434
Qy 117 rgSerArgProAsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValL 137
Db 25435 GAAGCAGGCCAAACTCAGGTGATGAGAAGTGCAAC--AGCCTGTCTGAAGGAAAGTGA 25492
Qy 137 ysLysAsnMetGlyGlyLysLysArgValSerSerGlySerSerGluArgGlySerL 157
Db 25493 AGAAGATCATGGGAGATGGAAGAGGGGGTGTCTTCGGGCTCTTCAGAGACAGGCTCCA 25552
Qy 157 ysSerProLeuLysArgAlaGlnGlnSerProArgLysArgGlyArgProProLysA 177
Db 25553 AATCCCTCTGAAAAGAAATCCAAGAGCAAAAGTCCCCAGAGCGGGGTCAAGCCCAAGA 25612
Qy 177 spGluLysAspLeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProM 197
Db 25613 ATGAGAAGAATGTCAACCATCGTGGAGTCCGGTACCTTGAAGTGATGATGCCGACCA 25672
Qy 197 etAlaAlaPheLysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisP 217
Db 25673 TGGCTGCATTTAAATGGCAGCCGACAGAGCGAGCTGTTAAAGATGCAGATCCTCAT 25732
Qy 237 ysLysLeuLysIleCysGluAsp 244
Db 25791 AGAAGTTGAAAATATGTGAAGAG 25813

RESULT 8

AL160172 231260 bp DNA linear HTG 13-JUN-2001
LOCUS Homo sapiens chromosome 1 clone RP11-6302 map q25.2-31.1, ***
DEFINITION SEQUENCING IN PROGRESS ***, 9 unordered pieces.
ACCESSION AL160172
VERSION AL160172.8 GI:14456182
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Plumb,B.
TITLE Direct Submission
JOURNAL Submitted (13-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Jun 14, 2001 this sequence version replaced gi:13373898.

COMMENT

----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bA6302
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M7815; 1% of reads

Sequencing vector: plasmid; L08752; 98% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 229290 bases at least Q40
Consensus quality: 229747 bases at least Q30
Consensus quality: 230088 bases at least Q20
Insert size: 230460; sum-of-contigs
Insert size: 235881; 6.2% error; agarose-fp
Quality coverage: 8.80x in Q20 bases; sum-of-contigs Quality
coverage: 9.32x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 32893: contig of 32893 bp in length
* 32894 32993: gap of 100 bp
* 32994 48692: contig of 15699 bp in length
* 48693 48792: gap of 100 bp
* 48793 64328: contig of 15536 bp in length
* 64329 64428: gap of 100 bp
* 64429 106773: contig of 42345 bp in length
* 106774 106873: gap of 100 bp
* 106874 120306: contig of 13433 bp in length
* 120307 120406: gap of 100 bp
* 120407 167991: contig of 47585 bp in length
* 167992 168091: gap of 100 bp
* 168092 176543: contig of 8452 bp in length
* 176544 208841: contig of 32198 bp in length
* 208842 208941: gap of 100 bp
* 208942 231260: contig of 22319 bp in length.

FEATURES

source
1. .231260
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="q25.2-31.1"
/clone="RP11-6302"
/clone_lib="RPC1-11.1"
1. .32893
/note="assembly_fragment:02273
clone_end:SP6
vector_side:left"
32994. .48692
/note="assembly_fragment:04073
fragment_chain:1"
48793. .64328
/note="assembly_fragment:03251
fragment_chain:1"
64429. .106773
/note="assembly_fragment:02028
fragment_chain:1"
106874. .120306
/note="assembly_fragment:01335
fragment_chain:1"
120407. .167991
/note="assembly_fragment:01561
fragment_chain:2"
168092. .176543
/note="assembly_fragment:04212
fragment_chain:2"
176644. .208841
/note="assembly_fragment:01546.0"
208942. .231260
/note="assembly_fragment:03725"
61279 a 52930 c 52964 g 63281 t 806 others

BASE COUNT
ORIGIN

Alignment Scores:		3.72e-53	Length:	231260
Pred. No.:		915.00	Matches:	204
Score:		63.04%	Conservative:	16
Best Local Similarity:		58.45%	Mismatches:	24
Query Match:		69.37%	Indels:	108
DB:		2	Gaps:	1
US-09-987-755-2 (1-249) x AL160172 (1-231260)				
QY	1	MetAlaAlaValSerLeuArgLeuGlyAspLeuValTrrpGlyLysLeuGlyArgTyrPro	20	
Db	66485	ATGGGGCTGTGAGTTCGCTCAGCAACTGTATGTGGGGAACCTCGGCTGGTATCCT	66544	
QY	21	ProTrrpProGlyLysIleValAsnProProLysAspLeuLysProArg-----	37	
Db	66545	CCTTCACCAGGAAGATCGTTAATCCACCAGGACCTTGAAGAACTCA-TGGGGCTGG	66603	
QY	37	-----	37	
Db	66604	CGCGGTGGCTCAGCCTGTGATCACAGCACTTTGGGAGGCCAAGGTGGCGGATCACGA	66663	
QY	37	-----	37	
Db	66664	GGTCAGGAGATTGAGACCATCTGGCTAACATGGTGAATCCCGTCTATATAAAATAC	66723	
QY	37	-----	37	
Db	66724	AAAAAAAATAATAGCCAGCGGTGGTGGCGGGCGTCTAGTCTCCTCAGGAGGCT	66783	
QY	37	-----	37	
Db	66784	GAGGCAGGAGAATAGCATGAACCCGGGAGGTGGAGCTTGCACTGAGCCGAGATCGCACCA	66843	
QY	37	-----	37	
Db	66844	CTGCACTCCAACCTGGGCAACAGAGTGAGACTCCATCTCAAAAAAAGAACCTCAT	66903	
QY	38	GlyLysLysCysPhePheValLys-PhePheGlyThrGluAspHisAlaTrrpLysVa	57	
Db	66904	GGAAAGAAATGCTTCTTTGTGAATAATTTTTTGAACATAGATCATCGCTGGATTAAAGT	66963	
QY	57	lGluGlnLeuLysProTrrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLy	77	
Db	66964	GAACAGCTGAAGCCATATCATGCTCATAAAGAGGAAATGATAAAGATTCAAGGGGTAA	67023	
QY	77	sArgPheGlnAlaValAspAlaValGluGluPheLeuArgArgAla-LysGlyLysA	97	
Db	67024	ACTATTCAGCAAGCGGTGGATGCCCAAGAGATTTCTCAGGAGAGCCCAAGGGAAAG	67083	
QY	97	spGlnThrSerSerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluA	117	
Db	67084	ATCAGACGTCATCCCACAATCTGTGTGATGACAAGAATCGGCTTAATTCAGTGAGGAGA	67143	
QY	117	rgSerArgProAsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValL	137	
Db	67144	GAAGCAGGCCAAACTCAGGTGATGAGAAGTGCAAAC--AGCCTGTCTGAAGGAAAAGTGA	67201	
QY	137	ysLysAsnMetGlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerL	157	
Db	67202	AGAAGATCATGGGAGATGGAAAGAAAGGGGTGTCTTCGGGCTCTTCAGAGACAGGCTCCA	67261	
QY	157	ysSerProLeuLysArgAlaGlnGluInSerProArgLysArgGlyArgProProLysA	177	
Db	67262	AATCCCTCTGAAAAGAAATCCAAAGAGCAAAGTCCCAGAAAGCGGGGTGAGCCCAAGA	67321	
QY	177	spGluLysAspLeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProM	197	
Db	67322	ATGAGAAGAATGTCAACCATCGTGGAGTCCCGTACCTTGAAGTGGATGATGGCGGACCCAG	67381	
QY	197	etAlaAlaPheLysTrrpGlnProThrAlaSerGluProValLysAspAlaAspProHisP	217	
Db	67382	TGGCTGCATTTAAATGGCAGCCGACCAAGAGCGAGCCTGTATAAGATGCAGATCCTCAT	67441	

Qy 217 heHisHisPheLeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrL 237
Db 67442 TCCATCATTTCTCTGTAAGCCAAACAGCCATC--TCTGTCTGTTACCAGGCAATCAAA 67499
Qy 237 ySLysLeuLysIleCysGluAsp 244
Db 67500 AGAAGTTGAAATATGTGAAGAG 67522
RESULT 9
BD059639/c
LOCUS BD059639 658 bp DNA linear PAT 27-AUG-2002
DEFINITION Secreted expressed sequence tags (sESTs).
ACCESSION BD059639
VERSION BD059639.1 GI:22605245
KEYWORDS JP 2001519666-A/1494.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 658)
AUTHORS Jacobs,K., Mccoy,J.M., Lavallie,E.R., Racie,L.A., Merberg,D.,
Treach,M., Spaulding,V. and Agostino,M.J.
TITLE Secreted expressed sequence tags (sESTs)
JOURNAL Patent: JP 2001519666-A 1494 23-OCT-2001;
GENETICS INSTITUTE INC
COMMENT PN JP 2001519666-A/1494
PD 23-OCT-2001
PF 10-APR-1998 JP 1998543068
PR 10-APR-1997 US 08/835913
PI KENNETH JACOBS,JOHN M MCCOY,EDWARD R LAVALLIE,LISA A RACIE, PI
DAVID MERBERG,
PI MAURICE TREACY,VIKKI SPAULDING,MICHAEL J AGOSTINO PC
C12N15/12,C12N5/10,C07K14/47,C12Q1/68,A61K38/17 CC Strandedness:
Double;
CC Topology: Linear;
FH Key Location/Qualifiers.
source 1. 658
/organism="Zea mays"
/mol_type="genomic DNA"
/db_xref="taxon:4577"
BASE COUNT 148 a 175 c 130 g 205 t
ORIGIN
Alignment Scores:
Pred. No.: 1.65e-33 Length: 658
Score: 595.50 Matches: 119
Percent Similarity: 97.54% Conservative: 0
Best Local Similarity: 97.54% Mismatches: 2
Query Match: 45.15% Indels: 1
DB: 6 Gaps: 1
US-09-987-755-2 (1-249) x BD059639 (1-658)
Qy 91 ArgArgAlaLysGlyLysAspGlnThrSerSerHisAsnSerSerAspAspLysAsnArg 110
Db 375 AGGAGAGCCAAAGGGAAAGACCAGACGTCATCCCAATTTCTGTATGACAAGAATCGA 316
Qy 111 ArgAsnSerSerGluArgSerArgProAsnSerGlyAspGluLysArgLysLeuSer 130
Db 315 CGTAATTCAGTGAGGAGAGAAAGTAGGCCAAACTCAGGTGATGAGAGCGCAAACTTAGC 256
Qy 131 LeuSerGluGlyLysValLysLysAsnMetGlyGluGlyLysArgValSerSerGly 150
Db 255 CTGTCTGAAGGGAAGGTGAAGAAGAACATGGGAGAGGAAGAAGAGGGGTGTTTCAGGC 196
Qy 151 SerSerGluArgGlySerLysSerProLeuLysArgAlaGlnGluGlnSerProArgLys 170
Db 195 TCTTCAGAGAGAGGCTCCAAATCCCTCTGAAAGAGAGCCCAAGAGCAAGTCCCCCGAAG 136
Qy 171 ArgGlyArgProProLysAspGluLysAspLeuThrIleProGluSerSerThrValLys 190

Db 135 CGGGTCGGCCCCCAAGGATGAGAAGGATCTCACCATCCCGGAGTCTAGTACCGTGAAG 76
Qy 191 GlyMetMetAlaGlyProMetAlaAlaPheLysTrpGlnProThrAla---SerGluPro 209
Db 75 GGGATGATGGCCGACCGATGGCCGGTTTAAATGGCAGCAACCGCGCTCTCTAGGCCA 16
Qy 210 Vallys 211
Db 15 TGAAG 10
RESULT 10
AC020663/c
LOCUS AC020663 154218 bp DNA linear PRI 08-JAN-2000
DEFINITION Homo sapiens chromosome 16 clone RPCI-11_127I20, complete sequence.
ACCESSION AC020663
VERSION AC020663.1 GI:6682593
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 154218)
AUTHORS Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,
Han,C. and Deaven,L.
TITLE Sequencing of Human Chromosome 16q12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 154218)
AUTHORS Ricke,D.O.
TITLE Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 154218)
AUTHORS Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,
McMurry,K., Han,C. and Deaven,L.
TITLE Direct Submission
JOURNAL Submitted (08-JAN-2000) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
FEATURES
source 1. 154218
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RPCI-11_127I20"
complement(41..330)
/rpt_family="Alu"
repeat_region 575..882
/rpt_family="Alu"
repeat_region 1414..1712
/rpt_family="Alu"
repeat_region 1595..1862
/rpt_family="Alu"
repeat_region 1871..2190
/rpt_family="Alu"
repeat_region 2176..2372
/rpt_family="L1"
repeat_region 2498..2795
/rpt_family="Alu"
repeat_region 3187..3504
/rpt_family="Alu"
repeat_region complement(3571..3890)
/rpt_family="Alu"
repeat_region complement(3859..4163)
/rpt_family="Alu"
repeat_region 4198..4503

```
repeat_region /rpt family="Alu"
4826. .5155
repeat_region /rpt family="Alu"
5749. .6045
misc_feature /rpt family="Alu"
complement(6848. .6983)
repeat_region /note="GRAIL 2 excellent exon, frame 1"
7473. .7776
repeat_region /rpt family="Alu"
7931. .8385
repeat_region /rpt family="Alu"
8808. .9139
repeat_region /rpt family="Alu"
9394. .9682
repeat_region /rpt family="Alu"
9515. .9812
repeat_region /rpt family="Alu"
9702. .9987
repeat_region /rpt family="Alu"
9965. .10332
repeat_region /rpt family="Alu"
complement(10403. .10835)
repeat_region /rpt family="Alu"
complement(10801. .10977)
repeat_region /rpt family="Alu"
11580. .11780
repeat_region /rpt family="MER42"
11803. .12101
repeat_region /rpt family="Alu"
12282. .12576
repeat_region /rpt family="Alu"
12643. .13073
repeat_region /rpt family="Alu"
12923. .13237
repeat_region /rpt family="Alu"
13420. .13725
repeat_region /rpt family="Alu"
13761. .14148
repeat_region /rpt family="Alu"
14072. .14371
repeat_region /rpt family="Alu"
complement(14401. .14696)
repeat_region /rpt family="Alu"
15061. .15366
repeat_region /rpt family="Alu"
15383. .15706
repeat_region /rpt family="Alu"
complement(16083. .16185)
repeat_region /rpt family="L1"
complement(16192. .16347)
repeat_region /rpt family="Alu"
16559. .16831
repeat_region /rpt family="Alu"
complement(18420. .18718)
repeat_region /rpt family="Alu"
complement(18518. .18917)
repeat_region /rpt family="Alu"
complement(18754. .19036)
misc_feature /rpt family="Alu"
complement(19456. .19559)
repeat_region /note="GRAIL 2 excellent exon, frame 1"
20239. .20397
repeat_region /rpt family="MER20"
complement(20808. .21104)
repeat_region /rpt family="Alu"
complement(22116. .22443)
misc_feature /rpt family="Alu"
complement(22789. .22948)
repeat_region /note="GRAIL 2 excellent exon, frame 1"
23346. .23646
repeat_region /rpt family="Alu"
complement(24046. .24341)
repeat_region /rpt family="Alu"

repeat_region 25205. .25461
/rpt family="Alu"
25605. .25862
/rpt family="L1"
25914. .25981
/rpt family="Alu"
complement(26044. .26355)
/rpt family="Alu"
28437. .28732
/rpt family="Alu"
30597. .30772
/note="GRAIL 2 excellent exon, frame 2"
31300. .31655
/rpt family="Alu"
32059. .32124
/rpt family="Alu"
complement(32302. .32492)
/rpt family="Alu"
32507. .32839
/rpt family="Alu"
33649. .34040
/rpt family="Alu"
complement(34202. .34278)
/rpt family="LTR37B"
34279. .34556
/rpt family="Alu"
35017. .35495
/rpt family="Alu"
complement(36639. .36991)
/rpt family="Alu"
37667. .37808
/note="GRAIL 2 excellent exon, frame 0"
complement(37920. .37961)
/rpt family="MIR"
38430. .38745
/rpt family="Alu"
38789. .38899
/rpt family="MER3"
complement(40939. .41089)
/rpt family="Alu"
42294. .42589
/rpt family="Alu"
complement(43093. .43396)
/rpt family="Alu"
complement(43274. .43695)
/rpt family="Alu"
44405. .44482
/rpt family="MLT1C"
44515. .44693
/rpt family="Alu"
45916. .46254
/rpt family="Alu"
complement(46300. .46347)
/note="GRAIL 2 excellent exon, frame 1"
46426. .46596
/rpt family="MER20"
47139. .47448
/rpt family="Alu"
47535. .47943
/rpt family="THE1"
complement(47939. .48377)
/rpt family="Alu"
48431. .48778
/rpt family="L1"
48877. .49260
/rpt family="Alu"
complement(49310. .49588)
/rpt family="Alu"

Alignment Scores:
Pred. No.: 8.72e-28 Length: 154218
Score: 551.00 Matches: 128
Percent Similarity: 40.92% Conservative: 5
Best Local Similarity: 39.38% Mismatches: 8
```


Query Match:	41.77%	Indels:	184
DB:	9	Gaps:	1
US-09-987-755-2 (1-249) x AC020663 (1-154218)			
Qy	53	AlaTrpIleLysValGluGlnLeuLysProTyrHisAlaHisLysGluGluMetIleLys	72
Db	125672	GCCTGGATCAAAGTGGAACAGCTGAAGCCATATCATGCTCATAAAGAGAAATGATAAAA	125613
Qy	73	IleAsnLysGlyLysArgPheGlnAlaValAspAlaValGluGluPheLeuArgArg	92
Db	125612	ATTAACAAGGGTAAACGATCCAGCAAGCGGTAGATGCTGTCGAAGAGTTCCTCAGGAGA	125553
Qy	93	AlaLysGlyLysAsp	97
Db	125552	GCCAAAGGAAAGACCAGGTGAGAGACACAATGTTCACTTTCTACCAATGCCAGACTGGT	125493
Qy	97		97
Db	125492	GATTTTGCATAGCAACTAACAGATTAAATGTACCCAAACGAAACCTTTGTTCTTCTG	125433
Qy	97		97
Db	125432	AAATGCTACTGGAGAAAGCTACTTCACTTAAGGTGTAGCGTTTGTGCTTATCAGTTTG	125373
Qy	97		97
Db	125372	AAACTCCATGCAGCTAACCTTTCTGAGTAGGTTGTTGCTTTAGGAAATGCTTGGAC	125313
Qy	97		97
Db	125312	CCAGGAGAAATATCTCAGCATTTGAGTGTAGCACAGGCCATTTCTTCCTTTCATAGCACA	125253
Qy	97		97
Db	125252	CAGATTAGGAGCATAGCCTCTGAAATACATATACATATACTGCCTAGATTAAATCCTTCTTCTA	125193
Qy	97		97
Db	125192	GTTTGGGACATAGGGAAGTGAAGTCTTTTGATATATATATTTTAAATCTGCTTCTCCCA	125133
Qy	97		97
Db	125132	ATGCTGGGACGACAGCAGAAATCTATTTCTAGGGTCACTGTGAGGAGACCACACAGAGTG	125073
Qy	97		97
Db	125072	CCTGTAGCAGCATGGCACCCTGTAAATGATGGTGGCGGAGCAGCTGTGGTG	125013
Qy	98		108
Db	125012	TAGTTATCATATTATTATTAAACTTTTCAGACGTCACTCCACAAATTTCTTCTGATGACAAG	124953
Qy	109	AsnArgArgAsnSerSerGluGluArgSerArgProAsnSerGlyAspGluLysArgLys	128
Db	124952	AATCGACGTAATTCAGTGGAGAGAGAACTAGGCCAACTCAGGTGATGAGAAAGCGCAA	124893
Qy	129	LeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGluGlyLysLysArgValSer	148
Db	124892	CTTAGCCTGTCTGAAGGAGAGGTGAAGAAACATGGGAGAGGAAAGAGAGGGTGTCT	124833
Qy	149	SerGlySerSerGluArgGlySerLysSerProLeuLysArgAlaGlnGluGlnSerPro	168
Db	124832	TCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAAAAGAGCCCCAAGCAAAAGTCCC	124773
Qy	169	ArgLysArgGlyArgProProLysAspGluLysAspLeuThrIleProGluSerSerThr	188
Db	124772	CGGAAGCGGGTCCGCCCCCAAGGATGAGAAGGTTTGTCTCTCTGACGCTTCCGGG	124713
Qy	189	VallLysGlyMetMet	193
Db	124712	ATTAGTGATGTCCTT	124698

RESULT 11	AC027687	169166 bp	DNA	linear	PRI 27-MAR-2003
LOCUS	Homo sapiens chromosome 16 clone RP11-61L4, complete sequence.				
DEFINITION	AC027687				
ACCESSION	AC027687.7	GI:29294025			
VERSION	HTG.				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 169166)				
AUTHORS	DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.				
TITLE	Direct Submission				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 169166)				
AUTHORS	DOE Joint Genome Institute.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-APR-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA				
REFERENCE	3 (bases 1 to 169166)				
AUTHORS	DOE Joint Genome Institute.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-MAR-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA				
REFERENCE	4 (bases 1 to 169166)				
AUTHORS	DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.				
TITLE	Direct Submission				
JOURNAL	Submitted (27-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA				
COMMENT	On Mar 27, 2003 this sequence version replaced gi:19033441. Draft Sequence Produced by DOE Joint Genome Institute				
FEATURES	source	1.169166			
		/organism="Homo sapiens"			
		/mol_type="genomic DNA"			
		/db_xref="taxon:9606"			
		/chromosome="16"			
		/clone="RP11-61L4"			
BASE COUNT	41124 a	42936 c	42004 g	43102 t	
ORIGIN					
Alignment Scores:		9.66e-28	Length:	169166	
Pred. No.:		551.00	Matches:	128	
Score:		40.92%	Conservative:	5	
Percent Similarity:		39.38%	Mismatches:	8	
Best Local Similarity:		41.77%	Indels:	184	
Query Match:		9	Gaps:	1	
DB:					
US-09-987-755-2 (1-249) x AC027687 (1-169166)					
Qy	53	AlaTrpIleLysValGluGlnLeuLysProTyrHisAlaHisLysGluGluMetIleLys	72		
Db	25492	GCCTGGATCAAAGTGGAACAGCTGAAGCCATATCATGCTCATAAAGAGAAATGATAAAA	25433		
Qy	73	IleAsnLysGlyLysArgPheGlnAlaValAspAlaValGluGluPheLeuArgArg	92		
Db	25432	ATTAACAAGGGTAAACGATTCAGCAAGCGGTAGATGCTGTCGAAGAGTTCCTCAGGAGA	25373		
Qy	93	AlaLysGlyLysAsp	97		
Db	25372	GCCAAAGGAAAGACCAGGTGAGAGACACAATGTTCACTTTCTACCAATGCCAGACTGGT	25313		
Qy	97		97		

Db	25312	GATTTTGCATAGCAACTAACAGATTAAATGTCACCCAAACGCAACCTTTGTCTTCTG	25253
QY	97	-----	97
Db	25252	AAATGCTACTGGAGAAAGCTACTTCAGTTCTAAGGTGTGACGGTTTGTCTTATCAGTTTG	25193
QY	97	-----	97
Db	25192	AAACTCCATGCAGCTAACCTTTCTGGAGTAGTGTGTTGCTTTTAGGAAAATTGCTTGGAC	25133
QY	97	-----	97
Db	25132	CCAGGAGAAATATCTCAGCATTTGAGTCTAGCACAGGCCATTTCTCCTTCATAGCACACA	25073
QY	97	-----	97
Db	25072	CAGATTAGGAGCATAGCCTCTGGAATACATATACTGCCTAGATTAAATCCTTTCTTCTTA	25013
QY	97	-----	97
Db	25012	GTTTGGGACATAGGAAAGTGACTTTTGTGATATATATATTTTAAATCTGCTTCTCCCA	24953
QY	97	-----	97
Db	24952	ATGCTGGGACGACAGCAGACAATCTATTCTTAGGTCATTGTGAGGAGACCACACAGAGTG	24893
QY	97	-----	97
Db	24892	CCTGTAGCAGCACATGGCACCTGTAAATGATGTGGCGATGGCGGACGAGCTGTGGTGG	24833
QY	98	-----GlnThrsSerHisAsnSerSerAspLys 108	
Db	24832	TAGTTATCATTTATTATTATTAACCTTCAGACGTCATCCCAATTTCTCTGATGACAAG	24773
QY	109	AsnArgArgAsnSerSerGluGluArgSerArgProAsnSerGlyAspGluLysArgLys 128	
Db	24772	AATCGACGTAATTCAGTGGAGAGAAGTAGGCCAAACTCAGGTGATGAGAAGCGCAA	24713
QY	129	LeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGluGlyLysLysArgValSer 148	
Db	24712	CTTAGCCTGTCTGAAGGAAGGTGAAGAAGAACATGGGAGAGAGGAAAGAGAGGGTGCT	24653
QY	149	SerGlySerSerGluArgGlySerLysSerProLeuLysArgAlaGlnGluGlnSerPro 168	
Db	24652	TCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAAAGAGAGCCCAAGAGCAAAAGTCCC	24593
QY	169	ArgLysArgGlyArgProProLysAspGluLysAspLeuThrIleProGluSerSerThr 188	
Db	24592	CGGAAGCGGGTTCGGCCCCCAAGGATGAGAAGTTGTTTCTCTCTGCAGCTCCGGG	24533
QY	189	ValLysGlyMetMet 193	
Db	24532	ATTAGTGATGTCCTT 24518	
RESULT 12			
AC113755/c			
LOCUS	AC113755	233786 bp	DNA linear HTG 19-NOV-2002
DEFINITION	Rattus norvegicus clone CH230-193C1, *** SEQUENCING IN PROGRESS		
	*** 3 unordered pieces.		
ACCESSION	AC113755		
VERSION	AC113755.5	GI:25072692	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.		
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
	Rattus.		
REFERENCE	1 (bases 1 to 233786)		
AUTHORS	Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,		
	Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,		
	AnyaIebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,		
	Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,		

Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederick,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,Z., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission
Unpublished
2 (bases 1 to 233786)
Worley,K.C.
Direct Submission
Submitted (05-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 233786)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23267215.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GSEQ

Center clone name: CH230-193C1

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 216369 bases at least Q40

Consensus quality: 218310 bases at least Q30

Consensus quality: 219420 bases at least Q20

Estimated insert size: 221075; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 11093: contig of 11093 bp in length
 * 11094 11193: gap of unknown length
 * 11194 232615: contig of 221422 bp in length
 * 232616 232715: gap of unknown length
 * 232716 233786: contig of 1071 bp in length.

FEATURES

source

1. .233786
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-193C1"

misc_feature

1. .1153
 /note="wgs end_extension
 clone_end:T7"
 978. .1446

misc_feature

/note="clone boundary
 clone_end:T7"
 site:
 end sequence:BH309210"

misc_feature

8020. .11093
 /note="wgs contig"

misc_feature

229660. .230825
 /note="wgs contig"

misc_feature

231327. .232615
 /note="wgs_contig"

BASE COUNT 55548 a 52773 c 53568 g 58339 t 13558 others
 ORIGIN

Alignment Scores:

Pred. No.: 5.46e-25 Length: 233786
 Score: 514.00 Matches: 123
 Percent Similarity: 44.56% Conservative: 4
 Best Local Similarity: 43.16% Mismatches: 7
 Query Match: 38.97% Indels: 151
 DB: 2 Gaps: 3

US-09-987-755-2 (1-249) x AC113755 (1-233786)

QY 39 LysLysCysPhePheValLysPheGlyThrGluAspHisAlaTrpIleLysValGlu 58
 |||||
 Db 204143 AAACCTCTGTTGTTT-----CCTACTAGCGCATGGATCAAGTGGAG 204102
 |||||
 QY 59 GlnLeuLysProTyrHisAlaHisLysGluMetIleLysIleAsnLysGlyLysArg 78
 |||||
 Db 204101 CAGCTGAAGCCTTACCATGCTCACAAGGAGAGATGATCAAGATCAACAGGCAAGCGG 204042
 |||||
 QY 79 PheGlnGlnAlaValAspAlaValGluPheLeuArgAlaLysGlyLysAsp--- 97
 |||||

Db 204041 TTCCAGCAAGCTGTGGATGCTGTGAGGAGTTTCCTCAGGAGAGCCAAAGGCAAGACCAG 203982
 QY 97 ----- 97
 Db 203981 GTGAGAGAGAGAGGGGCTGCAGTGGTCCCTCCAACATCAGGCCAGTGGTTTGCATGGTTC 203922
 QY 97 ----- 97
 Db 203921 TGAGAAGATGGTTTGTGTGACACACTCCTTTTCCGTGAAATGCCCTAGAAAGAGACACCTG 203862
 QY 97 ----- 97
 Db 203861 TGTCCAGAACTCTGCCCGTGGCTCCCTTTTCTAAGCATGTTGTTTGGGAGAGTGC 203802
 QY 97 ----- 97
 Db 203801 TCACCTTCACAGCACAGGCAAGAGCTTAGGAGTACAGCCACCTCTCACCTTTGGGGCTG 203742
 QY 97 ----- 97
 Db 203741 TAGACAAGCGACATGTTTCTTTTAAAGCTGCTCTTCCAGCACAAAGAGAGGTGAGC 203682
 QY 97 ----- 97
 Db 203681 AAGGATCCTGTTTCTCAGGGTGTATGAAGACAATTACAGAACTGCTGAGAGAACTGTG 203622
 QY 97 ----- 97
 Db 203621 TGGCCCATGTTGATGTCAGTGGTATAGTGGCTGTTGTAGTGGTTATCATTTATTGTT 203562
 QY 98 -----GlnThrSerSerHisAsnSerSerAspAspLysAsnArgAsnSerSerG1 115
 |||||
 Db 203561 GAACCTTCAGACATCATCTCACACTTCTGCTGATGACAAGATCGGCGTAATTCAGTGA 203502
 QY 115 uGluArgSerArgProAsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLy 135
 |||||
 Db 203501 GGAGAGAAGTAGGCCAAACTCAGGTGATGAGAAACGCAAGCTTAGCCTGCTGAAGGGAA 203442
 QY 135 sValLysLysAsnMetGlyGluGlyLysLysArgValSerSerGlySerSerGluArgG1 155
 |||||
 Db 203441 GGTGAAGAAGAACATCGGAGAGAGGAAGAGAGGGTGACTTCAGGCTCTGCAGACAGAG 203382
 QY 155 ySerLysSerProLeuLysArgAlaGlnGluGlnSerProArgLysArgGlyArgProPr 175
 |||||
 Db 203381 CTCCAAATGC---CTTAAAGAGCCCAAGAGCAAGTCCCGGAAGCGGGTCTGGCCCCC 203325
 QY 175 oLysAspGluLys 179
 |||||
 Db 203324 AAAGGATGAGAAG 203312

RESULT 13

AC108242
 LOCUS AC108242 239081 bp DNA linear HTG 13-NOV-2002
 DEFINITION Rattus norvegicus clone CH230-230G18, *** SEQUENCING IN PROGRESS
 ***, 2 unordered pieces.
 ACCESSION AC108242
 VERSION AC108242.5 GI:24942514
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 239081)
 Muzny,D,Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
 Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
 Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
 Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
 Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,

Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idiebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensheva,L., Loulsegged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelimeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Direct Submission
Unpublished
2 (bases 1 to 239081)
Worley,K.C.

Direct Submission

Submitted (27-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 239081)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 13, 2002 this sequence version replaced gi:23119074.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GJRI

Center clone name: CH230-230G18

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 225503 bases at least Q40

Consensus quality: 228480 bases at least Q30

Consensus quality: 230425 bases at least Q20

Estimated insert size: 233061; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 2 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 5876: contig of 5876 bp in length

* 5877 5976: gap of unknown length

* 5977 239081: contig of 233105 bp in length.

FEATURES

source

1. 239081

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-230G18"

1. 2515

/note="wgs_end_extension"

clone_end:T7"

4849..5674

/note="clone_boundary"

clone_end:T7"

site:

end sequence:RWBOF45TVB"

5977..7403

/note="wgs_contig"

complement(230992..231789)

/note="clone_boundary"

clone_end:Sp6

site:

end sequence:RWBOF45TVB"

BASE COUNT 61633 a 56132 c 55001 g 58761 t 7554 others

ORIGIN

Alignment Scores:

Pred. No.: 5.6e-25 Length: 239081

Score: 514.00 Matches: 123

Percent Similarity: 44.56% Conservative: 4

Best Local Similarity: 43.16% Mismatches: 7

Query Match: 38.97% Indels: 151

DB: 2 Gaps: 3

US-09-987-755-2 (1-249) x AC108242 (1-239081)

QY 39 LysLysCysPhePheValLysPheGlyThrGluAspHisAlaTrpIleLysValGlu 58
|||||

Db 93353 ANACTCTGTTGTTT-----CCTACTAGCGCATGGATCAAGTGGAG 93394
|||||

QY 59 GlnLeuLysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArg 78
|||||

Db 93395 CAGCTGAGCCTTACCATGCTCACAAGGAGGAGATGATCAAGATCAACACAGGCGG 93454
|||||

QY 79 PheGlnGlnAlaValAspAlaValGluGluPheLeuArgAlaLysGlyLysAsp--- 97
|||||

Db 93455 TTCCAGCAAGCTGTGGATGCTGCGAGGAGTTCCTCAGGAGAGCCCAAGAGACCAG 93514
|||||

QY 97 ----- 97

Db 93515 GTGAGAGAGAGAGGGGCTGCAGTGGTGTCCCTCCACATCAGGCCAGTGGTTTCATGGTTC 93574

QY 97 ----- 97
Db 93575 TGAGAAGATGTTGTGTGACAGACTCCTTTCCGTGAATGCCTAGAAGAGACACCTG 93634
QY 97 ----- 97
Db 93635 TGTCCAGAACTCTGCCCCGTGGCTCCCTTTTCTTAAGGCATGTTGTTTGGGAGGAGTCGC 93694
QY 97 ----- 97
Db 93695 TCACCTTCACAGCACAGCGCAAGAGCTTAGGAGTACAGCCACTCTCACCTTTGGGGCTG 93754
QY 97 ----- 97
Db 93755 TAGACAAGCGACATGTTTCTTTCTTTTAAAGCTGTCTTTCCAGCACAAAGAGAGGTGAGC 93814
QY 97 ----- 97
Db 93815 AAGGATCCTGTTCTCAGGGTGTATGAAGACAATTCACAGAACTGCTGAGAGAACTGTG 93874
QY 97 ----- 97
Db 93875 TGGCCCATGTTGATGGCAGTGGTGATAGTGGCTGTTGTAGTGGTTATCATTATTATTGTT 93934
QY 98 -----GlnThrSerSerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerG1 115
Db 93935 GAACTTTTCAGACATCATCTCACACTTCTGCTGATGACAAGAACTGGCGTAATTCAGTGA 93994
QY 115 uGluArgSerArgProAsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyL 135
Db 93995 GGAGAGAAGTAGGCCAACTCAGGTGATGAGAAACGCAAGCTTAGCCTGTCTGAAGGAA 94054
QY 135 sValLysLysAsnMetGlyGluGlyLysLysArgValSerSerGlySerSerGluArgG1 155
Db 94055 GGTGAAGAAGACATGGGAGAAGAAAGAGGGTGACTTCAGGCTCTGCAGACAGAGG 94114
QY 155 ySerLysSerProLeuLysArgAlaGlnGluInSerProArgLysArgGlyArgProPr 175
Db 94115 CTCCAAATGC---CTTAAAGAGCCCCAAGAGCAAGTCCCCGGAAGCGGGTGGCCCCC 94171
QY 175 oLysAspGluLys 179
Db 94172 AAAGGATGAGAAG 94184

RESULT 14
AC123492
LOCUS AC123492 243770 bp DNA linear HTG 12-OCT-2002
DEFINITION Rattus norvegicus clone CH230-100L13, WORKING DRAFT SEQUENCE.
ACCESSION AC123492
VERSION AC123492.3 GI:23666397
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 243770)
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,

Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokeleneh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 243770)
Worley,K.C.
Direct Submission
Submitted (29-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 243770)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (12-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 10, 2002 this sequence version replaced gi:21908419.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GXR
Center clone name: CH230-100L13
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 227678 bases at least Q40
Consensus quality: 229865 bases at least Q30
Consensus quality: 231324 bases at least Q20

Estimated insert size: 237575; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 243770: contig of 243770 bp in length.

FEATURES
source
1 .243770
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-100L13"
1
/note="clone_boundary
clone_end:T7
site:ECORI
end_sequence:BH305303"
complement(240521..241363)
/note="clone_boundary
clone_end:Sp6
site:ECORI
end_sequence:BH305306"

misc_feature
60464 a 56544 c 55414 g 59687 t 11661 others
ORIGIN

Alignment Scores:
Pred. No.: 5.72e-25 Length: 243770
Score: 514.00 Matches: 123
Percent Similarity: 44.56% Conservative: 4
Best Local Similarity: 43.16% Mismatches: 7
Query Match: 38.97% Indels: 151
DB: 2 Gaps: 3

US-09-987-755-2 (1-249) x AC123492 (1-243770)

QY 39 LysLysCysPhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGlu 58
|||||
DB 150572 AAACCTGTTGTTT-----CCTACTAGCGATGGATCAAGTGGAG 150613
|||||

QY 59 GlnLeuLysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArg 78
|||||
DB 150614 CAGGTGAAGCCTTACCATGCTCAAGAGGAGATGATCAAGATCAACAAGGCAAGCGG 150673
|||||

QY 79 PheGlnGlnAlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAsp--- 97
|||||
DB 150674 TTCCAGCAAGCTGTGGATGCTGTCGAGGAGTTCTCTCAGGAGAGCCAAAGGCAAGACCAG 150733
|||||

QY 97 ----- 97

DB 150734 GTGAGAGAGAGAGGGGCTGCAGTGGTCCCTCCCAACATCAGGCCAGTGGTTTGCATGGTTC 150793
|||||

QY 97 ----- 97

DB 150794 TGAGAAGATGGTTGTGTGTCAGACACTCCTTTTCGGTGAAATGCCTAGAGAGGACACCTG 150853
|||||

QY 97 ----- 97

DB 150854 TGTCAGAACTTGCCCGTGGCTCCCTTTTCTAAGGCATGTTGTTTGGGAGGAGTCGC 150913
|||||

QY 97 ----- 97

DB 150914 TCACCTTCACAGCAGGCAAGAGAGCTTAGGAGTACAGCCACCTCTCACCTTGGGGCTG 150973
|||||

QY 97 ----- 97

Db 150974 TAGACAAGCGACATGTTTTTTCTTTTAAAGCTGTCTTTCCAGCACAAAGAGAGGTGAGC 151033
QY 97 ----- 97

Db 151034 AAGGATCCTGTTTCTCAGGGTGCTATGAAGACAATTACAGAACTGCTGAGAGAACTGTG 151093
QY 97 ----- 97

Db 151094 TGGCCCATGTTGATGGCAGTGGTGATAGTGGCTGTTGTAGTGGTTATCATTTATTATTGTT 151153
QY 98 -----GlnThrSerSerHisAsnSerSerSerAspAspLysAsnArgArgAsnSerSerG1 115
|||||
Db 151154 GAACTTTCAGACATCATCTCACACTTCTGCTGATGACAAGAAATCGGCGTAATCCAGTGA 151213
QY 115 uGluArgSerArgProAsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLy 135
|||||
Db 151214 GGAGAGAAGTAGGCCAACTCAGGTGATGAGAAACGCAAGCTTAGCCTGTCTGAAGGGAA 151273
QY 135 sVallLysAsnMetGlyGluGlyLysLysArgValSerSerGlySerSerGluArgG1 155
|||||
Db 151274 GGTGAAGAAGAACATGGGAGAAAGAAAGAGAGGTGACTTCAGGCTCTGCAGACAGAGG 151333
QY 155 ySerLysSerProLeuLysArgAlaGlnGlnSerProArgLysArgGlyArgProPr 175
|||||

Db 151334 CTCCAAATGC---CTTAAAGAGCCCAAGAGCAAAAGTCCCGAAGCGGGTCTGGCCCCC 151390
QY 175 oLysAspGluLys 179
|||||

Db 151391 AAAGGATGAGAAG 151403

RESULT 15
AC127246/c
LOCUS AC127246 181475 bp DNA linear ROD 01-MAR-2003
DEFINITION Mus musculus chromosome 16 clone RP24-424L20, complete sequence.
ACCESSION AC127246
VERSION AC127246.3 GI:28626893
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 181475)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 181475)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 181475)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 181475)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Mar 1, 2003 this sequence version replaced gi:28015415.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:<http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BB0424L20
----- Location/Qualifiers -----
1. .181475

FEATURES
source

/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="16"
/clone="RP24-424L20"

BASE COUNT 46692 a 42460 c 43716 g 48607 t
ORIGIN

Alignment Scores:
Pred. No.: 1.64e-24 Length: 181475
Score: 505.50 Matches: 126
Percent Similarity: 44.41% Conservative: 5
Best Local Similarity: 42.71% Mismatches: 12
Query Match: 38.32% Indels: 153
DB: 10 Gaps: 4

US-09-987-755-2 (1-249) x AC127246 (1-181475)

Qy	43	PheValLysPhePheGlyThrGluAspHisAlaTirPileLysValGluGlnLeuLysPro	62
Db	38693	TTTCTGTGTTGTTTCTACTAGT-----GCCTGGATCAAAGTGAACAGCTAAAGCCT	38640
Qy	63	TyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnAla	82
Db	38639	TACCATGCTCACAGGAGGAGATGATAAGATTAAACAGGTAACCGTTCCAGCAAGCT	38580
Qy	83	ValAspAlaValGluPheLeuArgArgAlaLysGlyLysAspGln-----	98
Db	38579	GTGGATGCTGTGAAGAGTTCTCAGGAGAGCCAAAGGGAAGACCA-GGTGAGAGAGGC	38521
Qy	98	-----	98
Db	38520	TGCAGTCGTTCCCTCCCAACATCGGCTCAATGATTTTGGCATGGTTACTAGCAGATGGTGT	38461
Qy	98	-----	98
Db	38460	GTGTCATAGACAGGCTCCTTTCCCATGAGATTTCCTAGCAAGAAACAGCTGTGTCCAG	38401
Qy	98	-----	98
Db	38400	AACTCTATGTCTGGCTCACTTTTCTGATCATGCATGTTGTTTTCGGAAGAGTAGCCC	38341
Qy	98	-----	98
Db	38340	AGCTTCACAAATACAGGAAGAGATCTTAGGAATTACAGCCACCTTTGGGGCTGTAGACAAGT	38281
Qy	98	-----	98
Db	38280	GAGGTGATTTTTTTTTTTTAAAGCTGTCTTTCCAGCACAAACAGAGGTGAGCAAGGATG	38221
Qy	98	-----	98
Db	38220	CTATTTCTCAGGGTGGCTATGAAGACAGTTTCACAGAACTGTGAGAGAGAGAGAACTG	38161
Qy	98	-----	98
Db	38160	TGTGGCCCATGTTGATGGCAGTGGTGTAGTGGCTGTGTAGTGGTTATCATATTATTG	38101
Qy	99	-----ThrSerSerHisAsnSerSerAspLysAsnArgArgAsnSerSer	114
Db	38100	TTGAACCTTTCAGACATCATCCACACTTCTGCTGATGACAGAAATCGGCGTAATTCAGT	38041
Qy	115	GluGluArgSerArgProAsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGly	134
Db	38040	GAGGAGAGAACTAGGCCAAACTCAGGTGATGAGAAACGCAAGCTTAGCCTGTCTGAAGGG	37981
Qy	135	LysValLysLysAsnMetGlyGluGlyLysLysArgValSerSerGlySerSerGluArg	154
Db	37980	AAGGTGAAGAAAGAACATGGGAGAAGGAAAGAGAGGGTGACTTCAGGCTCTGCAGACAGA	37921
Qy	155	GlySerLysSerProLeuLysArgAlaGlnGluGlnSerProArgLysArgGlyArgPro	174
Db	37920	GCCTCCAAATGC---CTTAAAGAGAGCCCAAGAGCAAAAGTCCCCGGAAGCGGGGTCCGCC	37864

Qy	175	ProLysAspGluLys---AspLeuThrIleProGluSerSerThr	188
Db	37863	CCAAAGGATGAGAAAGGTTTGTCTTACTGCCGTAGCTTCTGGGACT	37819

Search completed: December 8, 2003, 12:46:49
Job time : 3978 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 8, 2003, 10:41:11 ; Search time 325 Seconds
(without alignments)
2546.388 Million cell updates/sec

Title: US-09-987-755-2
Perfect score: 1319
Sequence: 1 MAAVSLRLGDLVWGKLGKRYP.....VCYQAITKKLKICEDLLLP 249

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2201672 seqs, 1661799599 residues

Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09987755/runat_08122003_103436_18336/app_query.fasta_1.391
-DB=Published_Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09987755@cgn 1 1 221 @runat_08122003_103436_18336
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description

1	1319	100.0	3202	11	US-09-987-755-1	Sequence 1, Appli
2	1319	100.0	3256	15	US-10-103-313-190	Sequence 190, App
3	1314	99.6	3172	15	US-10-103-313-33	Sequence 33, Appl
4	1291	97.9	3727	13	US-10-067-482-1	Sequence 1, Appli
5	1278	96.9	3707	15	US-10-103-313-133	Sequence 133, App
6	1263	95.8	810	15	US-10-103-313-256	Sequence 256, App
7	193.5	14.7	1898	9	US-09-768-826-22	Sequence 22, Appl
8	193.5	14.7	3805	13	US-10-247-671-8	Sequence 8, Appli
9	192	14.6	2920	11	US-09-866-050A-541	Sequence 541, App
10	192	14.6	2920	15	US-10-152-661-541	Sequence 541, App
11	186.5	14.1	2133	9	US-09-925-302-68	Sequence 68, Appl
12	185.5	14.1	723	15	US-10-207-791-1	Sequence 1, Appli
13	185.5	14.1	2376	9	US-09-938-885A-4	Sequence 4, Appli
14	185.5	14.1	2376	15	US-10-207-791-3	Sequence 3, Appli
15	179.5	13.6	869	9	US-09-938-885A-2	Sequence 2, Appli
16	177.5	13.5	1919	9	US-09-925-301-145	Sequence 145, App
17	166	12.6	1297	9	US-09-768-826-24	Sequence 24, Appl
18	166	12.6	2272	11	US-09-946-374-307	Sequence 307, App
19	166	12.6	2272	13	US-10-015-387A-307	Sequence 307, App
20	166	12.6	2272	13	US-10-006-130A-307	Sequence 307, App
21	166	12.6	2272	13	US-10-199-672-345	Sequence 345, App
22	166	12.6	2272	13	US-10-006-172A-307	Sequence 307, App
23	166	12.6	2272	13	US-10-187-749-345	Sequence 345, App
24	166	12.6	2272	13	US-10-194-457-345	Sequence 345, App
25	166	12.6	2272	13	US-10-184-642-345	Sequence 345, App
26	166	12.6	2272	13	US-10-196-747-345	Sequence 345, App
27	166	12.6	2272	13	US-10-015-392A-307	Sequence 307, App
28	166	12.6	2272	13	US-10-017-253A-307	Sequence 307, App
29	166	12.6	2272	13	US-10-173-689-345	Sequence 345, App
30	166	12.6	2272	13	US-10-173-690-345	Sequence 345, App
31	166	12.6	2272	13	US-10-173-691-345	Sequence 345, App
32	166	12.6	2272	13	US-10-173-692-345	Sequence 345, App
33	166	12.6	2272	13	US-10-173-694-345	Sequence 345, App
34	166	12.6	2272	13	US-10-173-698-345	Sequence 345, App
35	166	12.6	2272	13	US-10-173-699-345	Sequence 345, App
36	166	12.6	2272	13	US-10-173-707-345	Sequence 345, App
37	166	12.6	2272	13	US-10-174-569-345	Sequence 345, App
38	166	12.6	2272	13	US-10-174-583-345	Sequence 345, App
39	166	12.6	2272	13	US-10-174-587-345	Sequence 345, App
40	166	12.6	2272	13	US-10-174-589-345	Sequence 345, App
41	166	12.6	2272	13	US-10-174-591-345	Sequence 345, App
42	166	12.6	2272	13	US-10-175-736-345	Sequence 345, App
43	166	12.6	2272	13	US-10-175-742-345	Sequence 345, App
44	166	12.6	2272	13	US-10-175-744-345	Sequence 345, App
45	166	12.6	2272	13	US-10-175-745-345	Sequence 345, App

ALIGNMENTS

RESULT 1
US-09-987-755-1
; Sequence 1, Application US/09987755
; Publication No. US20030022312A1
; GENERAL INFORMATION:
; APPLICANT: Kunsch et al
; TITLE OF INVENTION: Human Hepatoma-Derived Growth Factor-2
; FILE REFERENCE: PFI98D1C1
; CURRENT APPLICATION NUMBER: US/09/987,755
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 09/263,625
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: 08/464,600
; PRIOR FILING DATE: 1995-06-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 3202
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (34)..(780)
; OTHER INFORMATION:

US-09-987-755-1

Alignment Scores: 1.65e-134 Length: 3202
Pred. No.: 1319.00 Matches: 249
Score: 1319.00
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 11 Gaps: 0

US-09-987-755-2 (1-249) x US-09-987-755-1 (1-3202)

QY 1 MetAlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
Db 34 ATGGCGGCTGTAGTCTCGGCTCGGACTTGGTGTGGGGAACCTCGGCCGATATCCT 93
QY 21 ProTrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLys 40
Db 94 CCTTGGCCAGGAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAGAAA 153
QY 41 CysPhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu 60
Db 154 TGCTTCTTTGTGAAATTTTGGAAACAGAGATCATGCTGGATCAAAAGTGGACAGCTG 213
QY 61 LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln 80
Db 214 AAGCCATATCATGCTCATAAAGAGGAAATGATAAAATTAACAGGGTAAACGATTCAG 273
QY 81 GlnAlaValAspAlaValGluGluPheLeuArgAlaLysGlyLysAspGlnThrSer 100
Db 274 CAAGCGGTAGATGCTGCGAAGAGTTCCTCAGGAGAGCCAAAGGGAAGACAGACGTCA 333
QY 101 SerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro 120
Db 334 TCCACAAATTTCTTGATGACAAAGATCGACGTAATTCAGTGAGGAGAGAGTAGGCCA 393
QY 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet 140
Db 394 AACTCAGGTGATGAGAAAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAACAATG 453
QY 141 GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu 160
Db 454 GGAGAAGGAAGAAAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTG 513
QY 161 LysArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp 180
Db 514 AAAAGAGCCCAAGCAAAAGTCCCGGAAGCGGGTTCGGCCCCCAAGAGGATGAGAAGGAT 573
QY 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
Db 574 CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCGTTT 633
QY 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
Db 634 AAATGGCAGCCCAACCGCAAGCGAGCTGTTAAAGATGCAGATCCTCATTTCCATCATTTT 693
QY 221 LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys 240
Db 694 CTGCTAAGCCCAACAGAGAGCCAGCTGTCTGTATACAGGCAATCACGAAGAAGTTGAAA 753
QY 241 IleCysGluAspLeuLeuProArg 249
Db 754 ATATGTGAAGACCTCTCTTCTTCCTAGG 780

RESULT 2

US-10-103-313-190
; Sequence 190, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ207C1
; CURRENT APPLICATION NUMBER: US/10/103,313

; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 190
; LENGTH: 3256
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3220)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-103-313-190

Alignment Scores: 1.69e-134 Length: 3256
Pred. No.: 1319.00 Matches: 249
Score: 1319.00
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-09-987-755-2 (1-249) x US-10-103-313-190 (1-3256)

QY 1 MetAlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
Db 34 ATGGCGGCTGTAGTCTCGGCTCGGACTTGGTGTGGGGAACCTCGGCCGATATCCT 93
QY 21 ProTrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLys 40
Db 94 CCTTGGCCAGGAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAGAAA 153
QY 41 CysPhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu 60
Db 154 TGCTTCTTTGTGAAATTTTGGAAACAGAGATCATGCTGGATCAAAGTGGAAACAGCTG 213
QY 61 LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln 80
Db 214 AAGCCATATCATGCTCATAAAGAGGAAATGATAAAATTAACAGGGTAAACGATTCAG 273
QY 81 GlnAlaValAspAlaValGluGluPheLeuArgAlaLysGlyLysAspGlnThrSer 100
Db 274 CAAGCGGTAGATGCTGCGAAGAGTTCCTCAGGAGAGCCAAAGGGAAGACAGACGTCA 333
QY 101 SerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro 120
Db 334 TCCACAAATTTCTTGATGACAAAGATTCGACGTAATTCAGTGAGGAGAGAGTAGGCCA 393
QY 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet 140
Db 394 AACTCAGGTGATGAGAAAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAACAATG 453
QY 141 GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu 160
Db 454 GGAGAAGGAAGAAAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTG 513
QY 161 LysArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp 180
Db 514 AAAAGAGCCCAAGCAAAAGTCCCGGAAGCGGGTTCGGCCCCCAAGAGGATGAGAAGGAT 573
QY 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
Db 574 CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCGTTT 633
QY 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
Db 634 AAATGGCAGCCCAACCGCAAGCGAGCTGTTAAAGATGCAGATCCTCATTTCCATCATTTT 693
QY 221 LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys 240
Db 694 CTGCTAAGCCCAACAGAGAGCCAGCTGTCTGTATACAGGCAATCACGAAGAAGTTGAAA 753
QY 241 IleCysGluAspLeuLeuProArg 249

```
|||||
Db 754 ATATGTGAAGACCTCCTTCTCCTAGG 780
|||||
RESULT 3
US-10-103-313-33
; Sequence 33, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJZ07C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 3172
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-103-313-33
Alignment Scores:
Pred. No.: 5.77e-134 Length: 3172
Score: 1314.00 Matches: 248
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.62% Indels: 0
DB: 15 Gaps: 0
US-09-987-755-2 (1-249) x US-10-103-313-33 (1-3172)
Qy 2 AlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProPro 21
|||||
Db 9 GCGGCTGTGAGTCTGCGGCTCGCGACTTGGTGTGGGGAACCTCGGCCGATATCCTCT 68
|||||
Qy 22 TrpProGlyLysIleValAsnProProLysAspLeuLysProArgGlyLysLysCys 41
|||||
Db 69 TGGCCAGGAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAGAAATGC 128
|||||
Qy 42 PhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLys 61
|||||
Db 129 TTCTTTTGAAATTTTGGAAACAGAGATCATGCTCGATCAAAGTGGAAACAGCTGAAG 188
|||||
Qy 62 ProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGln 81
|||||
Db 189 CCATATCATGCTCATAAAGAGGAATGATAAAATTAACAGGGTAACGATTCAGCAA 248
|||||
Qy 82 AlaValAspAlaValGluPheLeuArgAlaLysGlyLysAspGlnThrSerSer 101
|||||
Db 249 GCGGTAGATGCTGTCGAAGAGTTCTCAGGAGAGCCAAAGGAAAGACCAGACGTCATCC 308
|||||
Qy 102 HisAsnSerSerAspAspLysAsnArgAsnSerSerGluGluArgSerArgProAsn 121
|||||
Db 309 CACAATTCCTCTGATGACAGAATCGACGTAATTCAGTGAGGAGAGAGTAGGCCAAAC 368
|||||
Qy 122 SerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGly 141
|||||
Db 369 TCAGGTGATGAGAAGCGCAACTTAGCCTGCTTGAAGGGAAGGTGAAGAAACATGGGA 428
|||||
Qy 142 GluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLys 161
|||||
Db 429 GAAGGAAAGAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAA 488
|||||
Qy 162 ArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAspLeu 181
|||||
Db 489 AGAGCCCAAGAGCAAGTCCCGAAGCGGGTTCGGCCCCCAAGGATGAGAGGATCTC 548
|||||
Qy 182 ThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLys 201
|||||
Db 549 ACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGCGCGACCGATGGCCGCGTTTAA 608
|||||
Qy 202 TrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPheLeu 221
|||||
|||||
Db 609 TGGCAGCCAAACCGCAGCGAGCCTGTTAAAGATGCAGATCCTCAITTCATCTCTG 668
|||||
Qy 222 LeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLysIle 241
|||||
Db 669 CTAAGCCAAACAGAGAGCCAGCTGTCTGTACCAGCAATCACGAGAAGTTGAAATA 728
|||||
Qy 242 CysGluAspLeuLeuLeuProArg 249
|||||
Db 729 TGTGAAGACCTCCTTCTCCTAGG 752
|||||
RESULT 4
US-10-067-482-1
; Sequence 1, Application US/10067482
; Publication No. US20030148407A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc.
; TITLE OF INVENTION: Angiogenesis Dehydrogenase Gene
; FILE REFERENCE: IU 102 R1
; CURRENT APPLICATION NUMBER: US/10/067,482
; CURRENT FILING DATE: 2002-02-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 3727
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (28)..(1689)
; OTHER INFORMATION:
US-10-067-482-1
Alignment Scores:
Pred. No.: 2.37e-131 Length: 3727
Score: 1291.00 Matches: 243
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 97.88% Indels: 0
DB: 13 Gaps: 0
US-09-987-755-2 (1-249) x US-10-067-482-1 (1-3727)
Qy 1 MetAlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
|||||
Db 28 ATGGCGGTGTGAGTCTGCGGCTCGCGACTTGGTGTGGGGAACCTCGGCCGATATCCT 87
|||||
Qy 21 ProTrpProGlyLysIleValAsnProProLysAspLeuLysProArgGlyLysLys 40
|||||
Db 88 CCTTGGCCAGGAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAGAAA 147
|||||
Qy 41 CysPhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu 60
|||||
Db 148 TGCTTCTTTGTGAAATTTTGGAAACAGAGATCATGCTGGATCAAAGTGGAAACAGCTG 207
|||||
Qy 61 LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln 80
|||||
Db 208 AAGCCATATCATGCTCATAAAGAGGAATGATAAAATTAACAGGGTAAACGATTCCAG 267
|||||
Qy 81 GlnAlaValAspAlaValGluPheLeuArgAlaLysGlyLysAspGlnThrSer 100
|||||
Db 268 CAAGCGGTAGATGCTGCGAAGAGTTCTCAGGAGAGCCAAAGGAAAGACCAGCGTCA 327
|||||
Qy 101 SerHisAsnSerSerAspAspLysAsnArgAsnSerSerGluGluArgSerArgPro 120
|||||
Db 328 TCCCACATTTCTTCTGATGACAAGAATCGACGTAATTCAGTGAGGAGAGAGTAGGCCA 387
|||||
Qy 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet 140
|||||
Db 388 AACTCAGGTGATGAGAGCGCAACTTAGCCTGTCTGAAGGGAAGGTGAAGAACATG 447
|||||
Qy 141 GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu 160
|||||
```

```
Db 448 GGAGAAGGAAAGAGGGGTGCTTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTG 507
Qy 161 LysArgAlaGlnGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp 180
Db 508 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGGTCCGCCCCCAAGAGATGAGAAGGAT 567
Qy 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
Db 568 CTCACCATCCCGAGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCGTTT 627
Qy 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
Db 628 AAATGGCAGCCCAACCGCAAGCGAGCCTGTTAAAGATGCAGATCCTCATTTCCATCATTTT 687
Qy 221 LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys 240
Db 688 CTGCTAAGCCCAACAGAGAAGCCAGCTGTCTGTATCCAGGCAATCACGAAGAAGTTGAAA 747
Qy 241 IleCysGluAsp 244
Db 748 ATATGTGAAGAG 759

RESULT 5
US-10-103-313-133
; Sequence 133, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJZ07C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 3707
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-103-313-133

Alignment Scores:
Pred. No.: 6.28e-130 Length: 3707
Score: 1278.00 Matches: 240
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 96.89% Indels: 0
DB: 15 Gaps: 0

US-09-987-755-2 (1-249) x US-10-103-313-133 (1-3707)
Qy 4 ValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProProTrpPro 23
Db 20 GTGAGTCTGCGGCTCGCGACTTGGTGTGGGGAAACTCGGCCGATATCCTCTTGGCCA 79
Qy 24 GlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLysCysPhePhe 43
Db 80 GGAAAGATTGTTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAATGCTTCTTT 139
Qy 44 ValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGlnLeuLysProTyr 63
Db 140 GTGAAATTTTTTGGAAACAGAAAGATCATGCTGGATCAAAAGTGGAAACAGCTGAAGCCATAT 199
Qy 64 HisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnAlaVal 83
Db 200 CATGCTCATAAAGAGGAAATGATAAAATTAACAAGGGTAAACGATTCCAGCAAGCGGTA 259
Qy 84 AspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSerSerHisAsn 103
Db 260 GATGCTGTGAAAGAGTTCTCTCAGGAGAGCCAAAGGAAAGACCAGACGTCATCCCAAT 319
Qy 104 SerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsnSerGly 123
```

```
Db 320 TCTTCTGATGACAAGAATCGACGTAATTCAGTTCAGGAGAGAGTAGGCCAAACTCAGGT 379
Qy 124 AspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGluGly 143
Db 380 GATGAGAAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAGAACATGGGAGAAGGA 439
Qy 144 LysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArgAla 163
Db 440 AAGAAAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTGAAAGAGCC 499
Qy 164 GlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAspLeuThrIle 183
Db 500 CAAGAGCAAAAGTCCCGGAAGCGGGGTCCGCCCCCAAGAGATGAGAAGGATCTCACCATC 559
Qy 184 ProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLysTrpGln 203
Db 560 CCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCGTTTAAATGGCAG 619
Qy 204 ProThrAlaSerGluProValLysAspAlaAspProHisPheHisPheLeuLeuSer 223
Db 620 CCAACCGCAAGCGAGCCTGTAAAGATGCAGATCCTCATTTCCATCATTTCTGCTAAGC 679
Qy 224 GlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLysIleCysGlu 243
Db 680 CAAACAGAGAAGCCAGCTGTCTGTATCCAGGCAATCACGAAGAAGTTGAAATATGTGAA 739
Qy 244 Asp 244
Db 740 GAG 742

RESULT 6
US-10-103-313-256
; Sequence 256, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJZ07C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 256
; LENGTH: 810
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (12)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (43)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-103-313-256

Alignment Scores:
Pred. No.: 3.9e-129 Length: 810
Score: 1263.00 Matches: 238
Percent Similarity: 99.17% Conservative: 1
Best Local Similarity: 98.76% Mismatches: 2
Query Match: 95.75% Indels: 0
DB: 15 Gaps: 0

US-09-987-755-2 (1-249) x US-10-103-313-256 (1-810)
Qy 4 ValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProProTrpPro 23
Db 2 GTGAGTCTGCGGCTCGCGACTTGGTGTGGGGAAACTCGGCGATATCCTCTTGGCCA 61
Qy 24 GlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLysCysPhePhe 43
Db 62 GGAAAGATTGTTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAATGCTTCTTT 121
```


QY 44 VallysPheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysProTyr 63
Db 122 GTGAAATTTTGGACAGAGATCATGCTGGATCAAGTGAACAGCTGAAGCCATAT 181
QY 64 HisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnAlaVal 83
Db 182 CATGCTCATAAAGAGGAATGATAAAATTAACAAGGGTAAACGATTCCACGAAGCGGTA 241
QY 84 AspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSerSerHisAsn 103
Db 242 GATGCTGTCAAGAGTTCTTCAGGAGAGCCAAAGGAAAGACACGATCCCAAT 301
QY 104 SerSerAspAspLysAsnArgAsnSerSerGluGluArgSerArgProAsnSerGly 123
Db 302 TCTTCTGATGACAAGAAATCGAGTAATTCAGTGAGGAGAGAAGTAGGCCAAACTCAGGT 361
QY 124 AspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGluGly 143
Db 362 GATGAGAAAGCCAAACTTAGCTGTCTGAAGGGAAGGTGAAGAAGAACATGGGAGAAAGGA 421
QY 144 LysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArgAla 163
Db 422 AAGAAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAAAAGAGCC 481
QY 164 GlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAspLeuThrIle 183
Db 482 CAAGAGCAAAGTYCCCGAAGCGGGGTGATGCCGGACCGATGGCCGCTTTAAATGGCAG 541
QY 184 ProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLysTrpGln 203
Db 542 CCGGAGTCTAGTACCGTGAAGGGGATGATGCCGGACCGATGGCCGCTTTAAATGGCAG 601
QY 204 ProThrAlaSerGluProValLysAspAlaAspProHisPheHisPheLeuLeuSer 223
Db 602 CCAACCGCAAGCGAGCCTGTAAAGATGCAGATCCTCATTTCCATCATTTCTGCTAAGC 661
QY 224 GlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLysIleCysGlu 243
Db 662 CAAACAGAGAAGCCAGCTGTCTGTATACCAGGCAATCACGAAGAAGTTGAATAATGTGAA 721
QY 244 Asp 244
Db 722 GAG 724

RESULT 7
US-09-768-826-22
; Sequence 22, Application US/09768826
; Patent No. US20020012966A1
; GENERAL INFORMATION:
; APPLICANT: Shi et al.
; TITLE OF INVENTION: 18 human secreted proteins
; FILE REFERENCE: PF512P1
; CURRENT APPLICATION NUMBER: US/09/768,826
; CURRENT FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US00/22350
; PRIOR FILING DATE: 2000-08-15
; PRIOR APPLICATION NUMBER: 60/148,759
; PRIOR FILING DATE: 1999-08-16
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1898
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-768-826-22

Alignment Scores:
Pred. No.: 2.45e-11 Length: 1898
Score: 193.50 Matches: 67
Percent Similarity: 45.26% Conservative: 38
Best Local Similarity: 28.88% Mismatches: 101
Query Match: 14.67% Indels: 27

DB: 9 Gaps: 7
US-09-987-755-2 (1-249) x US-09-768-826-22 (1-1898)
QY 7 ArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProProTrpProGlyLysIle 26
Db 177 AAAGCGGCGACCTGTCTTCGCCAAGATGAAGGGTACCCGCACTGGCGGCCCGGATT 236
QY 27 ValAsnProProLysAspLeuLysLysProArgGlyLysLysCysPheValLysPhe 46
Db 237 GATGAACCTCCAGAGGGCGCTGTGAAGCTCCAGCAACAAG---TATCCTATCTTCTTT 293
QY 47 PheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysProTyrHisAlaHis 66
Db 294 TTTGGCACCCCATGAAACTGCATTCTAGGTCCCAAGACCTTTTCCATATATAAGGAGTAC 353
QY 67 LysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnAlaValAspAlaVal 86
Db 354 AAAGACAAGTTTGGAAAGTCAAAACCAACGGAAGGATTTAACGAAGGATTGTGGGAATA 413
QY 87 Glu-----GluPheLeuArgArgAlaLysGlyLysAspGlnThrSerSer 101
Db 414 GAAATAAACCAGGAGTAAAGTTTACTGGCTACCAGGCAATTTCAGCAACAGAGCTCTTCA 473
QY 102 HisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsn 121
Db 474 GAAACTGAGGAGAGAGTGAATACTGCAGATGCAAGCAGTGAG-----GAA 521
QY 122 SerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLys---LysAsnMet 140
Db 522 GAAGTGATAGAGTAGAGAA-----GATGGAAGGCAAAAGAAAGAAATGAA 569
QY 141 GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLys-SerProLe 160
Db 570 AAAGCAGGCTCAAAACGGAAGTCAATATCTTCAAGAAATCCTCTAAACAGTCCC-- 627
QY 160 uLysArgAlaGlnGlnSerProArgLysArgGlyArgProProLysAspGluLysAs 180
Db 628 -GGAATCTCCAGGAGATGAAGATGACAAAGACTGCAAGAGAGGAGGAAACAAAGCAGC 686
QY 180 pLeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPh 200
Db 687 TCTGAGGTGGAGATGCCGGCAACGACACAAAGAAACACAACTT----- 729
QY 200 eLysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPh 220
Db 730 -----CAGACTTCAGAAACCCAGTGAGGGACCTAACTACCATATGAATGCTG 779
QY 220 eLeuLeuSerGlnThrGluLysProAlaValCys 231
Db 780 CATATTAAAGAGAAACCAC-AAGAAGGTATATGT 812

RESULT 8
US-10-247-671-8
; Sequence 8, Application US/10247671
; Publication No. US20030194721A1
; GENERAL INFORMATION:
; APPLICANT: Mikita, Thomas
; APPLICANT: Shiffman, Dov
; APPLICANT: Porter, Gordon, J.
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS
; FILE REFERENCE: PA-0050 US
; CURRENT APPLICATION NUMBER: US/10/247,671
; CURRENT FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: 60/323,784
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 186
; SOFTWARE: PERL Program
; SEQ ID NO 8
; LENGTH: 3805
; TYPE: DNA
; ORGANISM: Homo sapiens

; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030194721A1 2676869CB1
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 3788
; OTHER INFORMATION: a, t, c, g, or other
US-10-247-671-8

Alignment Scores:
Pred. No.: 6e-11 Length: 3805
Score: 193.50 Matches: 67
Percent Similarity: 45.26% Conservative: 38
Best Local Similarity: 28.88% Mismatches: 101
Query Match: 14.67% Indels: 27
DB: 13 Gaps: 7

US-09-987-755-2 (1-249) x US-10-247-671-8 (1-3805)

```
QY 7 ArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProProTrpProGlyLysIle 26
   :::::::::::::::::::::
Db 151 AAAGCGGCGACCTGGTCTTCGCCAAGATGAAGGCTACCCGCACTGGCGCGCGGATT 210
   :::::::::::::::::::::

QY 27 ValAsnProProLysAspLeuLysLysProArgGlyLysLysCysPhePheValLysPhe 46
   :::::::::::::::::::::
Db 211 GATGAACCTCCAGAGGCGCTGTGAAGCTCCAGCAACAAG---TATCCTATCTTCTTT 267
   :::::::::::::::::::::

QY 47 PheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysProTyrHisAlaHis 66
   :::::::::::::::::::::
Db 268 TTTGGCACCCATGAACCTGCAATTTCTAGGTCCCAAGACCTTTTCCATATAAGGAGTAC 327
   :::::::::::::::::::::

QY 67 LysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGlnAlaValAspAlaVal 86
   :::::::::::::::::::::
Db 328 AAAGACAAAGTTTGGAAAGTCAAAACGGAAGGATTTAACGAAGGATTGTGGAAATA 387
   :::::::::::::::::::::

QY 87 Glu-----GluPheLeuArgArgAlaLysGlyLysAspGlnThrSerSer 101
   :::::::::::::::::::::
Db 388 GAAATAAACCCAGGAGTAAAGTTTACTGGCTACAGGCAATTCAGCAACAGAGCTTTCA 447
   :::::::::::::::::::::

QY 102 HisAsnSerSerAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsn 121
   :::::::::::::::::::::
Db 448 GAAACTGAGGAGGAGGTGGAATACTACTGCAGTGCAGAGCAGTGA-----GAA 495
   :::::::::::::::::::::

QY 122 SerGlyAspGluLysArgLysLeuSerLeuSerGlyLysValLys-----LysAsnMet 140
   :::::::::::::::::::::
Db 496 GAAGGTGATAGATAGAAGAA-----GATGGAAGCAAGCAAGAAAGAAATGAA 543
   :::::::::::::::::::::

QY 141 GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLys-SerProLe 160
   :::::::::::::::::::::
Db 544 AAAGCAGGCTCAAAACGGGAAAAGTCATATACTTCAAGAAATCTCTAAACAGTCCC-- 601
   :::::::::::::::::::::

QY 160 uLysArgAlaGlnGlnSerProArgLysArgGlyArgProProLysAspGluLysAs 180
   :::::::::::::::::::::
Db 602 -GGAAATCTCCAGGAGATGAAGATGACAAAGACTGCAAGAAAGAGGAAACAAAGCAGC 660
   :::::::::::::::::::::

QY 180 pLeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPh 200
   :::::::::::::::::::::
Db 661 TCTGAGGTGGAGATGCGGGCAACGACACAAAGAAACAACTT----- 703
   :::::::::::::::::::::

QY 200 eLysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPh 220
   :::::::::::::::::::::
Db 704 -----CAGACTTGCAGAAACACAGTGAAGGACCTTAACCATATAATGAATGCTG 753
   :::::::::::::::::::::

QY 220 eLeuLeuSerGlnThrGluLysProAlaValCys 231
   :::::::::::::::::::::
Db 754 CATATTAAAGAAACCAC-AAGAAGGTTATATGT 786
   :::::::::::::::::::::
```

RESULT 9

US-09-866-050A-541
; Sequence 541, Application US/09866050A
; Publication No. US2003004071A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.

```
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011C4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 541
; LENGTH: 2920
; TYPE: DNA
; ORGANISM: Rat
US-09-866-050A-541

Alignment Scores:
Pred. No.: 6.23e-11 Length: 2920
Score: 192.00 Matches: 69
Percent Similarity: 43.41% Conservative: 43
Best Local Similarity: 26.74% Mismatches: 101
Query Match: 14.56% Indels: 45
DB: 11 Gaps: 9

US-09-987-755-2 (1-249) x US-09-866-050A-541 (1-2920)

QY 7 ArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProProTrpProGlyLysIle 26
   :::::::::::::::::::::
Db 311 AAAGCGGAGACCTGGTCTTCGCCAATAATGAAGGCTACCCGCACTGGCGCGCGGATT 370
   :::::::::::::::::::::

QY 27 ValAsnProProLysAspLeuLysLysProArgGlyLysLysCysPhePheValLysPhe 46
   :::::::::::::::::::::
Db 371 GATGAACCTCCAGAGGCGCGCTGAAGCCTCCAGCAACAAG---TACCCTATCTTCTTT 427
   :::::::::::::::::::::

QY 47 PheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysProTyrHisAlaHis 66
   :::::::::::::::::::::
Db 428 TTCGGAACCCATGAACCTGCATTTCTAGGTCTAAAGACCTTTTCCCATATAAGGAATAC 487
   :::::::::::::::::::::

QY 67 LysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGlnAlaValAspAlaVal 86
   :::::::::::::::::::::
Db 488 AAAGACAAAGTTTGGAAAGTCAAAACGGAAGGATTTAATGAAGGATTTATGGGAATTT 547
   :::::::::::::::::::::

QY 87 Glu-----GluPheLeuArgArgAlaLysGlyLysAspGlnThrSerSer 101
   :::::::::::::::::::::
Db 548 GAAATAATCCAGGAGTGAATTTACTGGGTACTCAGACAAATTCAGCAACAGAGCTCTTCA 607
   :::::::::::::::::::::

QY 102 HisAsnSerSerAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsn 121
   :::::::::::::::::::::
Db 608 GAAACTGAGGAGAGGAGGAGGAACTGCAGATGCAAGCAGTGA-----GAA 655
   :::::::::::::::::::::

QY 122 SerGlyAspGluLysArgLysLeuSerLeuSerGlyLysValLysValLysLysAsnMetGly 141
   :::::::::::::::::::::
Db 656 GAAGGTGAC-----AGAGTAGAAGATGGA 679
   :::::::::::::::::::::

QY 142 GluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLys 161
   :::::::::::::::::::::
Db 680 AAAGCAAGAGAGAAAGATGAAAAGGAGGCTCAAAACGGAAGAAAGTCTACACTTCAAAG 739
   :::::::::::::::::::::

QY 162 ArgAlaGlnGlnSerProArgLysArgGlyArgProProLysAspGluLysAspLeu 181
   :::::::::::::::::::::
Db 740 AAGTCTTCTAAACAGTCCCGGAAATCTCCAGGA-----GACGAAGATGACAAAGACTGC 793
   :::::::::::::::::::::

QY 182 ThrIleProGlu-----SerSerThrValLysGly-MetMetAlaGlyProMetAlaAl 199
   :::::::::::::::::::::
Db 794 AAAGAAGAGGAGAGCAAAAGCAGCTCTGAGGCGGAGATGCTGGCA-----ATGACACA 847
   :::::::::::::::::::::

QY 199 aPheLysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHis 219
   :::::::::::::::::::::
Db 848 AGAAACACGACTTCAGACTTGCAGAAAGCCGCTGAAGGACCTAACTACCGTAAATGAATG 907
   :::::::::::::::::::::

QY 219 sPheLeuLeuSerGlnThrGluLys----- 227
   :::::::::::::::::::::
```

```

Db      908 CTGCATATTGAGAGAAACACAAAGGTTAACTGTTGATTGCTGATTTCTTGATTG 967
Qy      228 -----ProAlaValCysTyrGlnAlaIleThrLysLysLeuLysLeuCys 242
Db      968 ATATGAACCAACAGTCTTT-----GTTGTCACCTGACAAAGCCCGAGTGTGT 1013

RESULT 10
US-10-152-661-541
; Sequence 541, Application US/10152661
; Publication No. US20030022835A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c5
; CURRENT APPLICATION NUMBER: US/10/152,661
; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 09/866,050
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/221,232
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: 60/206,650
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 09/312,283
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/NZ99/00051
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 09/188,930
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: 09/069,726
; PRIOR FILING DATE: 1998-04-29
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 541
; LENGTH: 2920
; TYPE: DNA
; ORGANISM: Rat
US-10-152-661-541

Alignment Scores:
Pred. No.:      6.23e-11      Length:      2920
Score:          192.00        Matches:      69
Percent Similarity: 43.41%    Conservative: 43
Best Local Similarity: 26.74%  Mismatches:    101
Query Match:     14.56%      Indels:       45
DB:              15          Gaps:           9

US-09-987-755-2 (1-249) x US-10-152-661-541 (1-2920)

Qy      7 ArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProProTrpProGlyLysIle 26
Db      311 AAAGCGGGAGACCTGGTCTTCGCCAAATGAAGGGCTACCGCACTGGCCCGCGGATT 370
Qy      27 ValAsnProProLysAspLeuLysLysProArgGlyLysLysCysPhePheValLysPhe 46
Db      371 GATGAACCTCCAGAGGGCGCGTGAAGCCTCCAGCAACAAAG---TACCCTATCTTCTTT 427
Qy      47 PheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysProTyrHisAlaHis 66
Db      428 TTCGGAACCCATGAACCTGCATTTCTAGGTCCTAAAGACCTTTTCCCATATAAGGAATAC 487
Qy      67 LysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGlnAlaValAspAlaVal 86
Db      488 AAAGACAAGTTTGGAAAGTCAAAACGGAAGGATTATTAATGAAGGATTATGGGAAATT 547
Qy      87 Glu-----GluPheLeuArgAlaLysGlyLysAspGlnThrSerSer 101
Db      109 GACGAGATGCTGAGGCTGCCGTGAAATCAACAGCCCAACAAA---TACCAAGTCTTTT 165

548 GAAAATAATCCAGGAGTGAATTTACTGGGTACCAGACAATTCAGCAACAGAGCTCTTCA 607
102 HisAsnSerSerAspAspLysAsnArgAsnSerSerGluGluArgSerArgProAsn 121
608 GAAACTGAGGAGAGAGGAGGAACTGCACTGCAGATGCAAGCAGTGAG-----GAA 655
122 SerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGly 141
656 GAAGGTGAC-----AGAGTAGAAGATGGA 679
142 GluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLys 161
680 AAAGCAAGAGAGAAAGAAATGAAAAGAGGAGCTCAAAACGGAAGAAAGTCTTCACTTCAAAG 739
162 ArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAspLeu 181
740 AAGTCTTCTAAACAGTCCCGGAAATCTCCAGGA-----GACGAAGATGACAAAGACTGC 793
182 ThrIleProGlu-----SerSerThrValLysGly-MetMetAlaGlyProMetAlaAl 199
794 AAAGAAGAGGAGAGAAACAAAGCAGCTCTGAGGGCGGAGATGCTGGCA-----ATGACACA 847
199 aPheLysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisH 219
848 AGAAACACGACTTCAGACTTGCAGAAAGCCGGTGAAGGGACCTAACTACCGTAATGAATG 907
219 sPheLeuLeuSerGlnThrGluLys----- 227
908 CTGCATATTGAGAGAAACCAACCAAGAGGTTAACTGTTGATTGCTGATTTCTTGATTG 967
228 -----ProAlaValCysTyrGlnAlaIleThrLysLysLeuLysIleCys 242
968 ATATGAACCAACAGTCTTT-----GTTGTCACCTGACAAAGCCCGAGTGTGT 1013

RESULT 11
US-09-925-302-68
; Sequence 68, Application US/09925302
; Patent No. US2002004941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 68
; LENGTH: 2133
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-302-68

Alignment Scores:
Pred. No.:      1.67e-10      Length:      2133
Score:          186.50        Matches:      64
Percent Similarity: 43.72%    Conservative: 30
Best Local Similarity: 29.77%  Mismatches:    100
Query Match:     14.14%      Indels:       21
DB:              9          Gaps:           8

US-09-987-755-2 (1-249) x US-09-925-302-68 (1-2133)

Qy      7 ArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProProTrpProGlyLysIle 26
Db      49 AAATGCGGGACCTGGTGTTCGCCAAGATGAAGGGCTACCCACACTGGCCGCCCGGATT 108
27 ValAsnProProLysAspLeuLysLysProArgGlyLysLysCysPhePheValLysPhe 46
109 GACGAGATGCTGAGGCTGCCGTGAAATCAACAGCCCAACAAA---TACCAAGTCTTTT 165
```

QY 47 PheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysProTyrHisAlaHis 66
Db 166 TTCGGACCCACGAGCGGCAATTCCTGGGCCCCAAAGACCTCTTCCCTTACGAGGAATCC 225
QY 67 LysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGlnAlaValAspAlaVal 86
Db 226 AAGGAGAAGTTTGGCAAGCCCCAACAAAGAGGAAAGGTTTCAGCGAGGGGCTGTGGGAGATC 285
QY 87 GluGluPheLeuArgAlaLysGlyLysAspGlnThrSerSerHisAsnSerSerAsp 106
Db 286 GAGAAC-----AACCTACTGTCAAGGCTTCCGGCTATCAGTCTCCTCCAG 330
QY 107 AspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsnSerGlyAspGluLys 126
Db 331 AAAAAGAGCTGTGTGGAAGAGCGCTGAACACAGAGCCCCGAAGCTGCAGAGGGTGACGGTGAT 390
QY 127 ArgLysLeuSerLeuSerGluGlyLysValLysAsnMetGlyGluGlyLysArg 146
Db 391 AAGAAG---GGGAATGCAGAGGGCAGCAGCAGCAG-----GAAGGGAAGCTGGTC 438
QY 147 ValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArgAlaGln----- 164
Db 439 ATTGATGAGCCAGCCAAAGGAGAAGAACAGAGAGCGTTGAAGAGGAGAGCAGGGGAC 498
QY 165 -----GluGlnSerProArg-----LysArgGlyArgProProLysAspGluLysAsp 180
Db 499 TTGCTGGAGGACTCTCTAAACGTCCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 555
QY 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
Db 556 -----AAGAGGCGAGCCACCTTGAGGTTGAGAGGCGCCCTTCTATGGAGGTGGAA 606
QY 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspPro 215
Db 607 AAG---AATAGCACCCCTCTGTAGCCCGGCTCTGCGCGGGGCGCT 648

RESULT 12

US-10-207-791-1
; Sequence 1, Application US/10207791
; Publication No. US20030120428A1
; GENERAL INFORMATION:
; APPLICANT: PharmaDesign, Inc; Masaki Mori
; TITLE OF INVENTION: A prediction method of the effect of radiotherapy for cancer pati
; FILE REFERENCE: PDP-0016
; CURRENT APPLICATION NUMBER: US/10/207,791
; CURRENT FILING DATE: 2002-07-31
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 723
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(723)
US-10-207-791-1

Alignment Scores:
Pred. No.: 5,32e-11 Length: 723
Score: 185.50 Matches: 64
Percent Similarity: 43.72% Conservative: 30
Best Local Similarity: 29.77% Mismatches: 100
Query Match: 14.06% Indels: 21
DB: 15 Gaps: 8

US-09-987-755-2 (1-249) x US-10-207-791-1 (1-723)

QY 7 ArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProTrpProGlyLysIle 26
Db 31 AAATGCGGGGACCTGGTGTTCGCCAAGATGAAGGCTACCCACACTGGCGCGCGGATT 90
QY 27 ValAsnProProLysAspLeuLysLysProArgGlyLysCysPhePheValLysPhe 46

Db 91 GACGAGATGCCTGAGGCTGCGGTGAAATCAACAGCCAAACAAA---TACCAAGTCTTTT 147
QY 47 PheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysProTyrHisAlaHis 66
Db 148 TTCGGGACCCACGAGCGGCAATTCCTGGGCCCCAAAGACCTCTTCCCTTACGAGGAATCC 207
QY 67 LysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGlnAlaValAspAlaVal 86
Db 208 AAGGAGAAGTTTGGCAAGCCCCAACAAAGAGGAAAGGTTTCAGCGAGGGGCTGTGGGAGATC 267
QY 87 GluGluPheLeuArgAlaLysGlyLysAspGlnThrSerSerHisAsnSerSerAsp 106
Db 268 GAGAAC-----AACCTACTGTCAAGGCTTCCGGCTATCAGTCTCCTCCAG 312
QY 107 AspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsnSerGlyAspGluLys 126
Db 313 AAAAAGAGCTGTGTGGAAGAGCGCTGAACACAGAGCCCCGAAGCTGCAGAGGGTGACGGTGAT 372
QY 127 ArgLysLeuSerLeuSerGluGlyLysValLysAsnMetGlyGluGlyLysArg 146
Db 373 AAGAAG---GGGAATGCAGAGGGCAGCAGCAGCAG-----GAAGGGAAGCTGGTC 420
QY 147 ValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArgAlaGln----- 164
Db 421 ATTGATGAGCCAGCCAAAGGAGAAGAACAGAGGAGCGTTGAAGAGGAGAGCAGGGGAC 480
QY 165 -----GluGlnSerProArg-----LysArgGlyArgProProLysAspGluLysAsp 180
Db 481 TTGCTGGAGGACTCTCTAAACGTCCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 537
QY 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
Db 538 -----AAGAGGCGAGCCACCTTGAGGTTGAGAGGCGCCCTTCTATGGAGGTGGAA 588
QY 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspPro 215
Db 589 AAG---AATAGCACCCCTCTGTAGCCCGGCTCTGCGCGGGGCGCT 630

RESULT 13

US-09-938-885A-4
; Sequence 4, Application US/09938885A
; Patent No. US20020090679A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Goli, Surya K.
; Murry, Lynn E.
; TITLE OF INVENTION: LUNG GROWTH FACTOR VARIANT
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/938,885A
; FILING DATE: 24-Aug-2001
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/760,745
; FILING DATE: 1996-12-07
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0169 US
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2376 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 598956
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-938-885A-4

Alignment Scores:
Pred. No.: 2.47e-10 Length: 2376
Score: 185.50 Matches: 64
Percent Similarity: 43.72% Conservative: 30
Best Local Similarity: 29.77% Mismatches: 100
Query Match: 14.06% Indels: 21
DB: 9 Gaps: 8

US-09-987-755-2 (1-249) x US-09-938-885A-4 (1-2376)

Qy 7 ArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProTrpProGlyLysile 26
Db 346 AAATGCGGGACCTGGTTCGCCAAGATGAAGGCTACCCACACTGGCGGCCCGGATT 405
Qy 27 ValAsnProProLysAspLeuLysLysProArgGlyLysLysCysPheValLysPhe 46
Db 406 GACGAGATGCTGAGGCTGCGGTGAATCAACAGCCAAACAAA---TACCAAGTCTTTT 462
Qy 47 PheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysProTyrHisAlaHis 66
Db 463 TTCGGGACCCACGAGACGGCATCTCTGGGCCCAAGACCTCTTCCCTTACGAGGAATCC 522
Qy 67 LysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGlnAlaValAlaVal 86
Db 523 AAGGAGAAGTTTGGCAAGCCCAACAGAGGAAGGTTTCAGCGAGGGGCTGTGGGAGATC 582
Qy 87 GluGluPheLeuArgAlaLysGlyLysAspGlnThrSerSerHisAsnSerSerAsp 106
Db 583 GAGAAC-----AACCTACTGTCAAGGCTTCCGGCTATCAGTCTCTCCAG 627
Qy 107 AspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsnSerGlyAspGluLys 126
Db 628 AAAAAGAGCTGTGTGAAGAGCCTGAACACAGAGCCGAGCTGCAGAGGTGACGGTGAT 687
Qy 127 ArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGluGlyLysArg 146
Db 688 AAGAAG---GGGAATGCAGAGGCGCAGCAGCAGCAG-----GAAGGGAAGCTGGTC 735
Qy 147 ValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArgAlaGln---- 164
Db 736 ATTGATGAGCCAGCCCAAGGAGAAGACGAGAAGGAGCGTTGAAGAGGAGCAGGGGAC 795
Qy 165 -----GluGlnSerProArg-----LysArgGlyArgProProLysAspGluLysAsp 180
Db 796 TTGCTGGAGGACTCTCTTAACGTCCCAAGGAGGCAGAAACCCCTGAAGGAGGAG--- 852
Qy 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
Db 853 -----AAGGAGGCAGCCACCTTGGAGGTTGAGAGGCCCTTCTCTATGAGGTGGAA 903
Qy 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspPro 215
Db 904 AAG---AATAGCACCCCTCTGAGCCCGGCTGTGGCCCGGGGCGCT 945

RESULT 14

US-10-207-791-3

; Sequence 3, Application US/10207791

; Publication No. US20030120428A1

; GENERAL INFORMATION:

; APPLICANT: PharmaDesign, Inc; Masaki Mori
; TITLE OF INVENTION: A prediction method of the effect of radiotherapy for cancer pati
; FILE REFERENCE: PDP-0016
; CURRENT APPLICATION NUMBER: US/10/207,791
; CURRENT FILING DATE: 2002-07-31
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2376
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (316)..(1038)
US-10-207-791-3

Alignment Scores:
Pred. No.: 2.47e-10 Length: 2376
Score: 185.50 Matches: 64
Percent Similarity: 43.72% Conservative: 30
Best Local Similarity: 29.77% Mismatches: 100
Query Match: 14.06% Indels: 21
DB: 15 Gaps: 8

US-09-987-755-2 (1-249) x US-10-207-791-3 (1-2376)

Qy 7 ArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProTrpProGlyLysile 26
Db 346 AAATGCGGGACCTGGTTCGCCAAGATGAAGGCTACCCACACTGGCGGCCCGGATT 405
Qy 27 ValAsnProProLysAspLeuLysLysProArgGlyLysLysCysPheValLysPhe 46
Db 406 GACGAGATGCTGAGGCTGCGGTGAATCAACAGCCAAACAAA---TACCAAGTCTTTT 462
Qy 47 PheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysProTyrHisAlaHis 66
Db 463 TTCGGGACCCACGAGACGGCATCTCTGGGCCCAAGACCTCTTCCCTTACGAGGAATCC 522
Qy 67 LysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGlnAlaValAlaVal 86
Db 523 AAGGAGAAGTTTGGCAAGCCCAACAGAGGAAGGTTTCAGCGAGGGGCTGTGGGAGATC 582
Qy 87 GluGluPheLeuArgAlaLysGlyLysAspGlnThrSerSerHisAsnSerSerAsp 106
Db 583 GAGAAC-----AACCTACTGTCAAGGCTTCCGGCTATCAGTCTCTCCAG 627
Qy 107 AspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsnSerGlyAspGluLys 126
Db 628 AAAAAGAGCTGTGTGAAGAGCCTGAACACAGAGCCGAGCTGCAGAGGTGACGGTGAT 687
Qy 127 ArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGluGlyLysArg 146
Db 688 AAGAAG---GGGAATGCAGAGGCGCAGCAGCAGCAG-----GAAGGGAAGCTGGTC 735
Qy 147 ValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArgAlaGln---- 164
Db 736 ATTGATGAGCCAGCCCAAGGAGAAGACGAGAAGGAGCGTTGAAGAGGAGCAGGGGAC 795
Qy 165 -----GluGlnSerProArg-----LysArgGlyArgProProLysAspGluLysAsp 180
Db 796 TTGCTGGAGGACTCTCTTAACGTCCCAAGGAGGCAGAAACCCCTGAAGGAGGAG--- 852
Qy 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
Db 853 -----AAGGAGGCAGCCACCTTGGAGGTTGAGAGGCCCTTCTCTATGAGGTGGAA 903
Qy 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspPro 215
Db 904 AAG---AATAGCACCCCTCTGAGCCCGGCTGTGGCCCGGGGCGCT 945

RESULT 15

US-09-938-885A-2

; Sequence 2, Application US/09938885A
; Patent No. US20020090679A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Goli, Surya K.
; Murry, Lynn E.
; TITLE OF INVENTION: LUNG GROWTH FACTOR VARIANT
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/938,885A
; FILING DATE: 24-Aug-2001
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/760,745
; FILING DATE: 1996-12-07
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0169 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 869 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: lungast01
; CLONE: 876242
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-938-885A-2

Alignment Scores:
Pred. No.: 3.07e-10 Length: 869
Score: 179.50 Matches: 66
Percent Similarity: 43.95% Conservative: 32
Best Local Similarity: 29.60% Mismatches: 100
Query Match: 13.61% Indels: 25
DB: 9 Gaps: 9

US-09-987-755-2 (1-249) x US-09-938-885A-2 (1-869)

QY 3 AlaValSerLeu-----ArgLeuGlyAspLeuValTrpGlyLysLeuGlyArg 18
Db 35 GCGTCAGCATGCCACACGCCCTTCAAGCCCGGGACTTGGTTCGCTAAGATGAAGGC 94
QY 19 TyrProProTrpProGlyLysIleValAsnProLysAspLeuLysLysProArgGly 38
Db 95 TACCCTCACTGGCCTGCCAGGATCGACGACATCGCGGATGGCGCGGTGAAGCCGCCACC 154
QY 39 LysLysCysPhePheValLysPheGlyThrGluAspHisAlaTrpIleLysValGlu 58
Db 155 AACAAAG---TACCCCATCTTTTCTTGGCACACACGAAACAGCTTCTTGGGCCCCCAA 211
QY 59 GlnLeuLysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArg 78
Db 212 GACCTCTTCCCTTACGAGGAATCCAAAGGAGAGTTTGGCAAGCCCAACAGAGAAAGGG 271

QY 79 PheGlnGlnAlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGln 98
Db 272 TTCAGCGAGGGGCTGTGGAGATCGAGAAC-----AACCTACTGTCAAG 316
QY 99 ThrSerSerHisAsnSerSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSer 118
Db 317 GCTTCCGGCTATCAGTCTCTCCAGAAAAGAGCTGTGTGAGAGCCTGAACCCAGAGCCC 376
QY 119 ArgProAsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLys 138
Db 377 GAAGCTGCAGAGGTGACGGGTGATAAGAAG---GGGAATGCAGAGGGCAGCAGCAGCAG 433
QY 139 AsnMetGlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSer 158
Db 434 -----GAAGGGAAGCTGCTCATTGATGAGCCAGCCAGGAGAGAAAGAACGAGAAAGGA 484
QY 159 ProLeuLysArgAlaGln-----GluGlnSerProArg-----LysArgGly 172
Db 485 GCGTTGAAGAGAGAGAGCAGGGGACTTGTCTGAGGACTCTCTCTAAACGTCCTCCAGGAGGCA 544
QY 173 ArgProProLysAspGluLysAspLeuThrIleProGluSerSerThrValLysGlyMet 192
Db 545 GAAACCCCTGAAGGAGAGGAG-----AAGGAGGAGAGCCACCTTGGAGGTTGAG 592
QY 193 MetAlaGlyProMetAlaAlaPheLysTrpGlnProThrAlaSerGluProValLysAsp 212
Db 593 AGGCCCCCTTCTATGGAGGTGGAAG---AATAGCACCCCTCTGTAGCCCCCGGCTCTGGC 649
QY 213 AlaAspPro 215
Db 650 CGGGGGCCT 658

Search completed: December 8, 2003, 12:48:20
Job time : 331 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 8, 2003, 10:34:41 ; Search time 279 Seconds
(without alignments)
2409.176 Million cell updates/sec.

Title: US-09-987-755-2
Perfect score: 1319
Sequence: 1 MAAVSLRLGLVWGKGRYP.....VCYQAITKKLKICEDLLLP 249

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09987755/runat_08122003_103435_18300/app_query.fasta_1.391
-DB=N Geneseq 19Jun03 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptp -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09987755 @CGN 1 1 312 @runat_08122003_103435_18300 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 19Jun03: *
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: *
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: *
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: *
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: *
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT: *
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: *
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT: *
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: *
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: *
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: *
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT: *
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT: *
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: *
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: *
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: *
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: *
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT: *
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: *
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: *
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT: *
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: *
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: *
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT: *
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: *
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1319	100.0	3202	18	AAT47520	Human hepatoma-der
2	1319	100.0	3256	22	AAS33262	DNA encoding human
3	1319	100.0	3256	22	AAS34946	CDNA encoding nove
4	1314	99.6	3172	22	AAS34789	CDNA encoding nove
5	1291	97.9	1939	25	AAL51565	Human nucleic acid
6	1291	97.9	3316	19	AAV44859	Clone AQ73_3 codin
7	1291	97.9	3316	22	AAF98456	Human cDNA clone A
8	1278	96.9	3707	22	AAS34889	CDNA encoding nove
9	1263	95.8	810	22	AAS33224	DNA encoding human
10	1263	95.8	810	22	AAS35012	CDNA encoding nove
c 11	591.5	44.8	658	20	AAV87516	EST clone BV27. H
12	504	38.2	2650	22	AAK70375	Human immune/haema
13	504	38.2	2650	22	AAK70376	Human immune/haema
c 14	444	33.7	307	20	AAV86389	EST clone AQ73. H
15	341	25.9	445	21	AAC21457	Human secreted pro
16	204.5	15.5	4487	24	ABS70439	Human bone remodel
17	194.5	14.7	3792	23	ABL03572	Drosophila melanog
18	193.5	14.7	1024	21	AAA13161	Human type II hepa
19	193.5	14.7	1898	22	AAF76854	Human secreted pro
20	193.5	14.7	1973	22	AAH14117	Human cDNA sequenc
21	193.5	14.7	3805	20	AAZ00048	HGFH2 gene. Homo
22	193	14.6	1713	23	ABL03573	Drosophila melanog
23	192	14.6	2920	24	ABL35012	Rat cDNA isolated
24	189	14.3	609	21	AAA65004	Human hepatoma-der
25	186.5	14.1	2133	21	AAF18049	Lung cancer associ
26	185.5	14.1	714	18	AAT88420	Mouse hepatoma der
27	185.5	14.1	723	16	AAQ79903	Human hepatoma der
28	185.5	14.1	1386	22	AAF59252	Human hHDGF nucleo
29	185.5	14.1	1563	18	AAV03867	Mouse hepatoma der
30	185.5	14.1	2376	16	AAQ79902	Human hepatoma der
31	185.5	14.1	2376	21	AAA40118	Human HDGFIN cDNA.
32	185.5	14.1	2376	22	ABA83087	Hepatoma-derived g
33	179.5	13.6	869	19	AAV39154	Lung growth factor
34	179	13.6	737	22	AAH05630	Human cDNA clone (
35	177.5	13.5	1919	21	AAC77751	Human cancer assoc
36	166	12.6	1297	22	AAF76856	Human secreted pro
37	166	12.6	2271	21	AAA37108	Human PRO1604 (UNQ
38	166	12.6	2272	22	AAS46097	Human DNA encoding
39	166	12.6	2272	22	AAF54420	DNA encoding prote
40	166	12.6	2272	25	ACA57855	Human PRO1604 cDNA
41	166	12.6	2272	25	ABX98325	Human cDNA encodin
42	166	12.6	2272	25	ABX98827	Novel human secret
43	166	12.6	2272	25	ACA05872	Human secreted/tra
44	166	12.6	2272	25	ABX97916	Human PRO polynucl
45	166	12.6	2272	25	ABX78700	Human PRO polynucl

ALIGNMENTS

RESULT 1
AAT47520
ID AAT47520 standard; cDNA; 3202 BP.
XX
AC AAT47520;
XX
DT 20-MAY-1997 (first entry)
XX
DE Human hepatoma-derived growth factor (HDGF-2) cDNA.
XX
KW Hepatoma-derived growth factor-2; HDGF-2; chemokine; wound healing;
KW vulnery; burn; ulcer; thrombosis; arteriosclerosis; gene therapy;
ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers

sequences may also be used as DNA probes in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate the presence of similar nucleic acid sequences in samples, and so which patients may be in need of restorative therapy. (II) may also be used as antigens in the production of antibodies and in assays to identify modulators (agonists and antagonists) of the expression and activity of the secreted proteins. The anti-(II) antibodies and antagonists may also be used to down regulate expression and activity of (II). The anti-(II) antibodies may also be used as diagnostic agents for detecting the presence of (II) in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The disorders include for example: immune/autoimmune diseases (e.g. HIV (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis and multiple sclerosis), cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of the breast or liver, Sezary syndrome and Gaucher's disease), neurological diseases (e.g. Alzheimer's disease, Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/cerebrovascular disorders (e.g. cardiac arrest, tachycardia, angina and thrombosis), infections caused by bacteria, viruses and fungi and ocular disorders (e.g. corneal infections). (I) and (II), agonists, antagonists and antibodies can also be used to promote wound healing, maintain organs before transplantation, and support cell culture of primary tissues. AAS33043-AAS33486 represent human secreted protein coding sequences, PCR primers, and related sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification but was obtained in electronic format directly from WIPO at: ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 3256 BP; 828 A; 835 C; 838 G; 753 T; 2 other;

Alignment Scores:

Pred. No.: 5, 78e-105 Length: 3256
Score: 1319.00 Matches: 249
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-09-987-755-2 (1-249) x AAS33262 (1-3256)

QY	1	MetAlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro	20
DB	34	ATGGCGGTGTGAGTCTGCGGCTCGCGACTTGGTGTGGGAAACTCGCGCATATCCT	93
QY	21	ProTrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLys	40
DB	94	CCTTGGCCAGGAAGATTGTTAATCCACCAAGGACTTGAAGAACTCGCGGAAGAAA	153
QY	41	CysPhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu	60
DB	154	TGCTTCTTTGTGAAATTTTTCGAACAGAGATCATGCTGGATCAAAAGTGAACAGCTG	213
QY	61	LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln	80
DB	214	AAGCCATATCATGCTCATAAAGAGGAATGATAAAATTAACAAGGTTAACGATTCCAG	273
QY	81	GlnAlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSer	100
DB	274	CAAGCGGTAGATGCTGTCTGAAGAGATTCTCTCAGGAGCCAAAGGAAAGACGATCA	333
QY	101	SerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro	120
DB	334	TCCCACAATTCTTCTGATGACAAGAATCGACGCTAATTCAGTCAGGAGAGTAGGCCA	393
QY	121	AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysAsnMet	140
DB	394	AACTCAGGTGATGAGAAGCGCAACTTAGCTGTCTGAAGGGAAGGTGAAGAACAATG	453
QY	141	GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu	160
DB	454	GGAGAAGGAAGAAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTG	513
QY	161	LysArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp	180

Db	514	AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTGGCCGCCCAAGATGAGAAGGAT	573
QY	181	LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe	200
Db	574	CTCACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCCGTTT	633
QY	201	LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe	220
Db	634	AAATGGCAGCCAACCGCAAGCGAGCTGTAAAGATGCAGATCCTCATTTCCATCATTT	693
QY	221	LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys	240
Db	694	CTGCTAAGCCAAACAGAGAAGCCAGCTGTCTGTATTACAGGCAATCACGAAGAAGTTG	753
QY	241	IleCysGluAspLeuLeuLeuProArg	249
Db	754	ATATGTGAAGACCTCCTTCTTCTTAGG	780
RESULT 3			
AAS34946			
ID	AAS34946 standard; cDNA; 3256 BP.		
AC	AAS34946;		
XX	04-DEC-2001 (first entry)		
DT	04-DEC-2001 (first entry)		
XX	cDNA encoding novel human neoplastic disease associated polypeptide #180.		
XX	Human; neoplastic disease associated polypeptide; cancer; gene therapy;		
KW	hyperproliferative disorder; neural disorder; immune system disorder;		
KW	muscular disorder; reproductive disorder; gastrointestinal disorder;		
KW	pulmonary disorder; cardiovascular disorder; renal disorder;		
KW	neuroprotective; cytostatic; anti inflammatory; vasotropic; ss.		
OS	Homo sapiens.		
XX	WO200155163-A1.		
PN	02-AUG-2001.		
PD	17-JAN-2001; 2001WO-US01358.		
XX	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	24-FEB-2000; 2000US-0184664.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0189874.		
PR	17-MAR-2000; 2000US-0190076.		
PR	18-APR-2000; 2000US-0198123.		
PR	19-MAY-2000; 2000US-0205515.		
PR	07-JUN-2000; 2000US-0209467.		
PR	28-JUN-2000; 2000US-0214886.		
PR	30-JUN-2000; 2000US-0215135.		
PR	07-JUL-2000; 2000US-0216647.		
PR	07-JUL-2000; 2000US-0216880.		
PR	11-JUL-2000; 2000US-0217487.		
PR	11-JUL-2000; 2000US-0217496.		
PR	14-JUL-2000; 2000US-0218290.		
PR	26-JUL-2000; 2000US-0220963.		
PR	26-JUL-2000; 2000US-0220964.		
PR	14-AUG-2000; 2000US-0224518.		
PR	14-AUG-2000; 2000US-0224519.		
PR	14-AUG-2000; 2000US-0225213.		
PR	14-AUG-2000; 2000US-0225214.		
PR	14-AUG-2000; 2000US-0225266.		
PR	14-AUG-2000; 2000US-0225267.		
PR	14-AUG-2000; 2000US-0225268.		
PR	14-AUG-2000; 2000US-0225270.		
PR	14-AUG-2000; 2000US-0225447.		
PR	14-AUG-2000; 2000US-0225757.		
PR	14-AUG-2000; 2000US-0225758.		
PR	14-AUG-2000; 2000US-0225759.		
PR	18-AUG-2000; 2000US-0226279.		

PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465558/50.

P-PSDB; AAU21747.

Novel polypeptides and polynucleotides useful as diagnostic reagents to
diagnose diseases or disorders associated with aberrant expression or
activity of polypeptides, and for treating cancers, rheumatoid
arthritis -

Claim 4; SEQ ID No 190; 687pp; English.

The present invention relates to the isolation of novel human neoplastic
disease associated polypeptides (AAU21568-AAU21851), and cDNA and DNA
sequences encoding for these polypeptides. The sequences of the
invention are useful in the diagnosis, treatment, prevention and/or
prognosis of disorders involving neoplastic disease such as
hyperproliferative disorders (e.g. leukaemia, bone cancer, bladder
cancer, brain stem glioma, adult liver cancer, childhood cerebellar
astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may
also be useful for treating other disorders such as neural disorders,
immune system disorders, muscular disorders, reproductive disorders,
gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
and renal disorders. The polynucleotide sequences of the invention are
also useful in gene therapy. AAS34767-AAS35050 represent cDNA sequences
encoding for the novel human neoplastic disease associated polypeptides
of the invention.

Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.

Sequence 3256 BP; 828 A; 835 C; 838 G; 753 T; 2 other;

Alignment Scores:

Pred. No.:	5.78e-105	Length:	3256
Score:	1319.00	Matches:	249
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0

Query Match:	100.00%	Indels:	0
DB:	22	Gaps:	0
US-09-987-755-2 (1-249) x AAS34946 (1-3256)			
QY	1	MetAlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro	20
Db	34	ATGGCGCTGTGAGTCTGGGCTCGGCGACTTGGTGTGGGGAAACTCGGCCGATATCCT	93
QY	21	ProTrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLys	40
Db	94	CCTTGGCCAGGAAGATGTTAATCCACCAAGGACTTGAAGAACTCGCGGAAGAAA	153
QY	41	CysPhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu	60
Db	154	TGCTTCTTTGTGAAATTTTGGACACAGAAGATCATGCTGGATCAAAGTGGACAGCTG	213
QY	61	LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln	80
Db	214	AAGCCATATCATGCTCATAAAGAGGAATGATAAAATTAACAAGGTAAACGATTCAG	273
QY	81	GlnAlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSer	100
Db	274	CAAGCGGTAGATGCTGTCCAAGAGTTCTCAGGAGAGCAAAAGGGAAGACCAAGCTCA	333
QY	101	SerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro	120
Db	334	TCCACAAATTTCTTGATGACAAGATCCACGTAATTCAGTGAGGAGAGAGTAGGCCA	393
QY	121	AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet	140
Db	394	AACTCAGTGATGAGAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTCAAGGAAGCATG	453
QY	141	GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu	160
Db	454	GGAGAAGGAAAGAGAGGTGTCTTCAGGCTCTTCAGAGAGAGGTCCAAATCCCTCTG	513
QY	161	LysArgAlaGlnGluInSerProArgLysArgGlyArgProProLysAspGluLysAsp	180
Db	514	AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGGATGAGAAGAT	573
QY	181	LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe	200
Db	574	CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGTTT	633
QY	201	LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe	220
Db	634	AAATGGCAGCCCAACCGCAAGCGAGCCTGTTAAAGATGAGATCCTCAITTCATATTC	693
QY	221	LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys	240
Db	694	CTGCTAAGCCAAACAGAGAAGCCAGCTGTCTGTACCAGGCAATCACCAAGAGTTGAA	753
QY	241	IleCysGluAspLeuLeuLeuProArg	249
Db	754	ATATGTGAAGACCTCCTTCTTCTTAGG	780
RESULT 4			
AAS34789			
ID	AAS34789 standard; cDNA; 3172 BP.		
XX			
AC	AAS34789;		
XX			
DT	04-DEC-2001 (first entry)		
XX			
DE	cDNA encoding novel human neoplastic disease associated polypeptide #23.		
XX			
KW	Human; neoplastic disease associated polypeptide; cancer; gene therapy;		
KW	hyperproliferative disorder; neural disorder; immune system disorder;		
KW	muscular disorder; reproductive disorder; gastrointestinal disorder;		
KW	pulmonary disorder; cardiovascular disorder; renal disorder;		
KW	neuroprotective; cytostatic; anti inflammatory; vasotropic; ss.		
XX			

OS	Homo sapiens.
XX	
PN	WO200155163-A1.
XX	
PD	02-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01358.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	14-JUL-2000; 2000US-0218290.
PR	26-JUL-2000; 2000US-0220963.
PR	26-JUL-2000; 2000US-0220964.
PR	14-AUG-2000; 2000US-0224518.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0225213.
PR	14-AUG-2000; 2000US-0225214.
PR	14-AUG-2000; 2000US-0225266.
PR	14-AUG-2000; 2000US-0225267.
PR	14-AUG-2000; 2000US-0225268.
PR	14-AUG-2000; 2000US-0225270.
PR	14-AUG-2000; 2000US-0225447.
PR	14-AUG-2000; 2000US-0225757.
PR	14-AUG-2000; 2000US-0225758.
PR	14-AUG-2000; 2000US-0225759.
PR	18-AUG-2000; 2000US-0226279.
PR	22-AUG-2000; 2000US-0226681.
PR	22-AUG-2000; 2000US-0226868.
PR	22-AUG-2000; 2000US-0227182.
PR	23-AUG-2000; 2000US-0227009.
PR	30-AUG-2000; 2000US-0228924.
PR	01-SEP-2000; 2000US-0229287.
PR	01-SEP-2000; 2000US-0229343.
PR	01-SEP-2000; 2000US-0229344.
PR	01-SEP-2000; 2000US-0229345.
PR	05-SEP-2000; 2000US-0229509.
PR	05-SEP-2000; 2000US-0229513.
PR	06-SEP-2000; 2000US-0230437.
PR	06-SEP-2000; 2000US-0230438.
PR	08-SEP-2000; 2000US-0231242.
PR	08-SEP-2000; 2000US-0231243.
PR	08-SEP-2000; 2000US-0231244.
PR	08-SEP-2000; 2000US-0231413.
PR	08-SEP-2000; 2000US-0231414.
PR	08-SEP-2000; 2000US-0232080.
PR	08-SEP-2000; 2000US-0232081.
PR	12-SEP-2000; 2000US-0231968.
PR	14-SEP-2000; 2000US-0232397.
PR	14-SEP-2000; 2000US-0232398.
PR	14-SEP-2000; 2000US-0232399.
PR	14-SEP-2000; 2000US-0232400.
PR	14-SEP-2000; 2000US-0232401.
PR	14-SEP-2000; 2000US-0233063.
PR	14-SEP-2000; 2000US-0233064.
PR	14-SEP-2000; 2000US-0233065.
PR	21-SEP-2000; 2000US-0234223.
PR	21-SEP-2000; 2000US-0234274.
PR	25-SEP-2000; 2000US-0234997.
PR	25-SEP-2000; 2000US-0234998.
PR	26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465558/50.
DR P-PSDB; AAU21590.
XX
PT Novel polypeptides and polynucleotides useful as diagnostic reagents to
PT diagnose diseases or disorders associated with aberrant expression or
PT activity of polypeptides, and for treating cancers, rheumatoid
PT arthritis -
XX
PS Claim 4; SEQ ID No 33; 687pp; English.
XX
CC The present invention relates to the isolation of novel human neoplastic
CC disease associated polypeptides (AAU21568-AAU21851), and cDNA and DNA
CC sequences encoding for these polypeptides. The sequences of the
CC invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of disorders involving neoplastic disease such as
CC hyperproliferative disorders (e.g. leukaemia, bone cancer, bladder
CC cancer, brain stem glioma, adult liver cancer, childhood cerebellar
CC astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may
CC also be useful for treating other disorders such as neural disorders,
CC immune system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC and renal disorders. The polynucleotide sequences of the invention are
CC also useful in gene therapy. AAS34767-AAS35050 represent cDNA sequences
CC encoding for the novel human neoplastic disease associated polypeptides
CC of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3172 BP; 814 A; 818 C; 808 G; 732 T; 0 other;

Alignment Scores:
Pred. No.: 1.52e-104 Length: 3172
Score: 1314.00 Matches: 248
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.62% Indels: 0
DB: 22 Gaps: 0

US-09-987-755-2 (1-249) x AAS34789 (1-3172)

QY 2 AlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProPro 21
Db 9 GCGGCTGTGAGTCTGCGGCTCGCGGACTTGGTGTGGGGGAAACTCGGCCGATATCCTCCT 68
QY 22 TrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLysCys 41
Db 69 TGGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAATGC 128
QY 42 PhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLys 61
Db 129 TTCTTTGTGAATTTTGGAAACAGAGATCATGCTGGATCAAAGTGGACAGCTGAAG 188
QY 62 ProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGln 81
Db 189 CCATATCATGCTCATAAAGAGAGAAATGATAAAAAATTACAGGGTAAACGATTCCAGCAA 248
QY 82 AlaValAspAlaValGluPheLeuArgAlaLysGlyLysAspGlnThrSerSer 101
Db 249 GCGGTAGATGCTGTGGAAGATTCTCTCAGGAGAGCCAAAGGGAAAGACCAGACTCATCC 308
QY 102 HisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsn 121
Db 309 CACAATTCTTCTGATGACAGAAATCGACGTAATTCAGTGAGGAGAGAGTAGGCCAAAC 368
QY 122 SerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGly 141
Db 369 TCAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGAAGGGGAGGTGAAGAGAACATGGGA 428
QY 142 GluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLys 161

Db 429 GAAGGAAAGAGAGGGTGTCTTCAGGCTCTCAGAGAGAGGCTCCAATCCCTCTGAAA 488
Qy 162 ArgAlaGlnGluSerProArgLysArgGlyArgProProLysAspGluLysAspLeu 181
Db 489 AGAGCCCAAGAGCAAAGTCCCGGAAGCGGGTGGGCCCCCAAGAGGATGAGAAGGATCTC 548
Qy 182 ThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLys 201
Db 549 ACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCGCGTTTAAA 608
Qy 202 TrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPheLeu 221
Db 609 TGGCAGCCCAACCGCAAGCGACCTGTTAAAGATGCAGATCTCATTTCCATCATTTCCCTG 668
Qy 222 LeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLysIle 241
Db 669 CTAAGCCCAACAGAGAGAGCGAGCTGTCTGTACCAGGCAATCACGAAGAAGTTGAAAATA 728
Qy 242 CysGluAspLeuLeuProArg 249
Db 729 TGTGAAGACCTCTCTTCTTCTTAGG 752

RESULT 5

AAL51565

ID AAL51565 standard; DNA; 1939 BP.

XX AC AAL51565;

XX DT 10-APR-2003 (first entry)

XX DE Human nucleic acid-associated protein coding sequence - SEQ ID No 48.

XX KW Human; gene; ds; nucleic acid-associated protein; NAAP; arteriosclerosis;
KW cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS;
KW cancer; developmental disorder; renal tubular acidosis; anaemia; asthma;
KW mental retardation; neurological disorder; Alzheimer's disease; epilepsy;
KW Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy;
KW Crohn's disease; transgenic animal; animal model.

OS Homo sapiens.

XX PN WO2003000864-A2.

XX PD 03-JAN-2003.

XX PF 20-JUN-2002; 2002WO-US21179.

XX PR 22-JUN-2001; 2001US-300518P.

XX PR 29-JUN-2001; 2001US-301787P.

XX PR 29-JUN-2001; 2001US-301792P.

XX PR 29-JUN-2001; 2001US-301892P.

XX PR 29-JUN-2001; 2001US-301893P.

XX PR 06-JUL-2001; 2001US-303405P.

XX PR 06-JUL-2001; 2001US-303442P.

XX PR 15-MAR-2002; 2002US-364438P.

XX PA (INCY-) INCYTE GENOMICS INC.

XX PI Gandhi AR, Swarnakar A, Hafalia AJA, Warren BA, Emerling BM;

XX PI Arvizu CS, Ison CH, Honchell CD, Lee EA, Yue H, Forsythe IJ;

XX PI Ramkumar J, Griffin JA, Yang J, Sanjanwala MM, Baughn MR;

XX PI Borowsky ML, Yao MG, Wallia NK, Bandman O, Lal PG, Becha SD;

XX PI Lee SY, Richardson TW, Elliott VS, Luo W, Tang YT, Zebarjadian Y;

XX PI Lu Y;

XX DR WPI; 2003-201420/19.

XX DR P-PSDB; AAO16415.

XX PT New nucleic acid-associated proteins and polynucleotides, useful for
PT diagnosing, treating or preventing cell proliferative (e.g. cancer),
PT neurological (e.g. epilepsy or Parkinson's disease), or autoimmune
PT disorders (e.g. AIDS) -

XX PS Claim 12; Page 285-286; 312pp; English.
XX CC The invention comprises the amino acid and coding sequences of human
CC nucleic acid-associated proteins (NAAP). The DNA and protein sequences of
CC the invention are useful for diagnosing, treating or preventing disorders
CC associated with aberrant expression of NAAP, such as: cell proliferative
CC disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis
CC or cancer); developmental disorders (e.g. renal tubular acidosis, anaemia
CC or mental retardation); neurological disorders (e.g. Alzheimer's disease,
CC Parkinson's disease or epilepsy); and autoimmune/inflammatory disorders
CC (e.g. AIDS, allergies, asthma or Crohn's disease). The DNA sequences of
CC the invention are useful for creating transgenic animals to model human
CC disease. The present DNA sequence encodes a human nucleic acid-associated
CC protein of the invention.

XX SQ Sequence 1939 BP; 504 A; 498 C; 548 G; 389 T; 0 other;

Alignment Scores:

Pred. No.: 8.41e-103 Length: 1939
Score: 1291.00 Matches: 243
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 97.88% Indels: 0
DB: 25 Gaps: 0

US-09-987-755-2 (1-249) x AAL51565 (1-1939)

Qy 1 MetAlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
Db 9 ATGGCGGTGTGAGTCTGGGCTCGGACTTGGTGTGGGGAACCTCGGCCGATATCCT 68
Qy 21 ProTrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLys 40
Db 69 CCTTGGCCAGAAAGATTGTTAATCCACCAAGGACTTCAAGAAACCTCGCGAAAGAAA 128
Qy 41 CysPhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu 60
Db 129 TGCTTCTTGTGAAATTTTGGAAACAGAGATCATGCTGGATCAAAGTGAACAGCTG 188
Qy 61 LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln 80
Db 189 AAGCCATATCATGCTCATAAAGAGGAATGATAAAATTAACAAGGGTAAACGATTCCAG 248
Qy 81 GlnAlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSer 100
Db 249 CAAGCGGTAGATGCTGTCAAAGAGTTCTCTCAGGAGAGCCAAAGGGAAAGACGACGTC 308
Qy 101 SerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro 120
Db 309 TCCCAACAATCTTCTGATGACAAGAATCGACGTAATTCAGTGAGGAGAGAAGTAGGCCA 368
Qy 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerGlySerGlyLysValLysAsnMet 140
Db 369 AACTCAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAGACATG 428
Qy 141 GlyGluGlyLysLysArgValSerSerGlySerSerGlySerGlySerLysSerProLeu 160
Db 429 GGAGAAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAATCCCTCTCTG 488
Qy 161 LysArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp 180
Db 489 AAAAGAGCCCAAGAGCAAAGTCCCGGAAGCGGGGTGCGGCCCCCAAGAGGATGAGAAGG 548
Qy 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
Db 549 CTCACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCCGCTTT 608
Qy 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
Db 609 AAATGGCAGCAACCGCAAGCGAGCTGTAAAGATGCAGATCCTCATTTCCATCATTTTC 668
Qy 221 LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys 240

Db 669 CTGCTAAGCCAAACAGAGAGCCAGCTGTCTGTACCAGGCAATCACGAAGAAGTTGAAA 728

QY 241 IleCysGluAsp 244
Db 729 ATATGTGAAGAG 740

RESULT 6
AAV44859
ID AAV44859 standard; cDNA; 3316 BP.
XX
AC AAV44859;

XX
DT 21-OCT-1998 (first entry)
DE Clone AQ73_3 coding sequence.

XX
KW Secreted protein; nutritional source; cell proliferation activity;
KW cell differentiation activity; immune stimulant; tissue growth activator;
KW haematopoiesis regulator; anti-inflammatory; tumour invasion suppressor;
KW tumour inhibitor; clone AQ73_3; ds.

OS Homo sapiens.

XX
FH Key Location/Qualifiers
FT CDS 7..1650
FT /*tag= a

XX PN WO9825962-A2.

XX PD 18-JUN-1998.

XX PF 12-DEC-1997; 97WO-US23224.

XX PR 11-DEC-1997; 97US-0989232.

XX PR 13-DEC-1996; 96US-0766263.

XX (GEMY) GENETICS INST INC.

XX PI Agostino MJ, Jacobs K, Lavallie ER, Mccoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;

XX DR WPI; 1998-362424/31.
XX DR P-PSDB; AAW69240.

XX PT New isolated polynucleotides - obtained from human adult testis,
PT human adult ovary, human adult brain and human adult heart cDNA
PT libraries

XX PS Claim 14; Page 63-65; 108pp; English.

XX
CC This sequence represents a polynucleotide of the invention, and encodes a
CC secreted protein. It was isolated from a human adult ovary cDNA library,
CC and is designated clone AQ73_3. The DNA sequences and encoded
CC polypeptides can be used as nutritional sources or supplements, or may
CC exhibit e.g. cytokine and cell proliferation/differentiation activity,
CC immune stimulating or suppressing activity, haematopoiesis regulating
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC activin/inhibin activity, chemostatic/chemokinetic activity,
CC cadherin/tumour invasion suppressor activity, tissue growth activity,
CC tumour inhibition activity or other activities.

XX SQ Sequence 3316 BP; 875 A; 813 C; 887 G; 741 T; 0 other;

Alignment Scores:
Pred. No.: 1.6e-102 Length: 3316
Score: 1291.00 Matches: 243
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 97.88% Indels: 0
DB: 19 Gaps: 0

US-09-987-755-2 (1-249) x AAV44859 (1-3316)

QY 1 MetAlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
Db 7 ATGGCGGCTGTGAGTCTGGCGCTCGGCGACTTGGTGTGGGGGAAACTCGGCCGATATCCT 66
QY 21 ProTrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLys 40
Db 67 CCTTGGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAA 126
QY 41 CysPhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu 60
Db 127 TGCTTCTTTGTGAAATTTTGTGAACAGAAAGATCATGCTCGATCAAAGTGAACAGCTG 186
QY 61 LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln 80
Db 187 AAGCCATATCATGCTCATAAAGAGGAAATGATAAAATTAAACAGGGTAAACGATTCAG 246
QY 81 GlnAlaValAspAlaValGluGluPheLeuArgAlaLysGlyLysAspGlnThrSer 100
Db 247 CAAGCGGTAGTGTCTCGAAGAGTTCTCTCAGGAGAGCCAAAGGGAAAGACCAGACGTCA 306
QY 101 SerHisAsnSerSerAspAspLysAsnArgAsnSerSerGluGluArgSerArgPro 120
Db 307 TCCACAAATTTCTTCTGATGACAAGAAATCGACGTAATTCAGTGAAGAGGAGAGTAGGCCA 366
QY 121 AsnSerGlyAspGluLysArgLysLeuSerSerLeuSerGluGlyLysValLysAsnMet 140
Db 367 AACTCAGGTGATGAGAGAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAACATG 426
QY 141 GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu 160
Db 427 GGAGAAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGGGCTCCAAATCCCCCTCTG 486
QY 161 LysArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp 180
Db 487 AAAAGAGCCCAAGAGCAAAGTCCCCGGAAGCGGGGTTCGGCCCCCAAAGGATGAGAAGGAT 546
QY 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
Db 547 CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCGGACCGATGGCGCGTTT 606
QY 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisHisPhe 220
Db 607 AAATGGCAGCCCAACCGCAAGCGAGCCTGTATAAGATGCAGATCCTCATTTCCATCATTTT 666
QY 221 LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys 240
Db 667 CTGCTAAGCCAAACAGAGAGCCAGCTGTCTGTATACCAGGCAATCACGAAGAAGTTGAAA 726
QY 241 IleCysGluAsp 244
Db 727 ATATGTGAAGAG 738
RESULT 7
ID AAF98456
XX AAF98456 standard; cDNA; 3316 BP.
AC AAF98456;
XX
DT 07-JUN-2001 (first entry)
XX
DE Human cDNA clone AQ73_3 sequence SEQ ID 136.
XX
KW Human; secreted protein; nutrient; cytokine modulator; proliferation;
KW differentiation; immune system modulator; tissue growth; chemotactic;
KW haemostatic; thrombolytic; anti-inflammatory; tumour inhibition; ss;
KW haematopoiesis.
OS Homo sapiens.
XX
XX WO200119988-A1.
XX PN
XX PD 22-MAR-2001.

XX 14-SEP-2000; 2000WO-US25135.
PF 17-SEP-1999; 99US-0398829.
PR (GEMY) GENETICS INST INC.
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Bowman MR, Spaulding V, Agostino MJ;
PI WPI; 2001-244801/25.
DR P-PSDB; AAB90720.
XX Isolated nucleic acids encoding polypeptides, useful for modulating
PT e.g. cytokine and cell proliferation/differentiation activity, the
PT immune system and hematopoiesis regulating activity -
XX Disclosure; Page 467-468; 557pp; English.
XX Human cDNA clones represented in AAF98374 - AAF98489 encode secreted
CC proteins AAB90667 - AAB90750. The cDNA clones are isolated from various
CC tissue types, and may be used in the prevention, treatment and diagnosis
CC of diseases associated with inappropriate protein expression. The
CC polypeptides and nucleic acids may be used as nutrients or to modulate
CC cytokine and cell proliferation/differentiation activity and may also be
CC involved in modulation of the immune system. The cDNA sequences,
CC proteins, their agonists and/or antagonists exhibit haematopoiesis
CC regulating activity; tissue growth activity; activin/inhibin activity;
CC chemotactic/chemokinetic activity; haemostatic and thrombolytic
CC activity; receptor/ligand activity; anti-inflammatory activity;
CC haematopoiesis activity; cadherin/tumour suppressor activity; and/or
CC tumour inhibition activity. Included in the invention are probes
CC represented in AAF98490 - AAF98572 which are specific for the cDNA clones
CC encoding the secreted proteins.
XX
SQ Sequence 3316 BP; 875 A; 813 C; 887 G; 741 T; 0 other;

Alignment Scores:
Pred. No.: 1.6e-102 Length: 3316
Score: 1291.00 Matches: 243
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 97.88% Indels: 0
DB: 22 Gaps: 0

US-09-987-755-2 (1-249) x AAF98456 (1-3316)

QY 1 MetAlaAlaValSerLeuArgLeuGlyAspLeuValTIpGlyLysLeuGlyArgTyrPro 20
Db 7 ATGGCGGCTGTGAGTCTCGCGCTCGGCGACTGGTGTGGGGAAACTCGGCCGATATCCT 66

QY 21 ProTIpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLys 40
Db 67 CCTTGGCCAGGAAGATTGTTAATCCACCAAGGACTTGAGAAACCTCGCGGAAGAAA 126

QY 41 CysPhePheValLysPheGlyThrGluAspHisAlaTIpIleLysValGluGlnLeu 60
Db 127 TGCTTCTTTGTGAAATTTTGGAAACAGAGATCATGCTCGGTGATCAAAAGTGGACAGCTG 186

QY 61 LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln 80
Db 187 AAGCCATATCATGCTCATAAAGAGGAATGATAAAATTAACAAGGGTAAACGATTCCAG 246

QY 81 GlnAlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSer 100
Db 247 CAAGCGGTAGATGCTGTGGAAGAGTTCTCTCAGGAGAGCCAAAGGGAAAGACGACGTCA 306

QY 101 SerHisAsnSerAspAspLysAsnArgAsnSerSerGluGluArgSerArgPro 120
Db 307 TCCCACAAATCTTCTGTATGACAAGAAATCGACGTAAATTCAGTGAGGAGAGTAGGCCA 366

QY 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet 140

Db 367 AACTCAGGTGATGAGAAGCGCAAACTTAGCTGTCTGAAGGGAAGGTGAAGAACAATG 426
QY 141 GlyGluGlyLysLysArgValSerSerGlySerGlySerGluArgGlySerLysSerProLeu 160
Db 427 GGAGAAGGAAGAAAGAGGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG 486
QY 161 LysArgAlaGlnGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp 180
Db 487 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGAGATGAGAAGAT 546
QY 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
Db 547 CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGCGCGACCGATGGCCGCTTT 606
QY 201 LysTIpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
Db 607 AAATGGCAGCCAACCGCAAGCGAGCTGTTAAAGATGAGATCCTCATTTCCATCATTTTC 666
QY 221 LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys 240
Db 667 CTGCTAAGCCCAACAGAGAAGCCAGCTGTCTGTACCAGGCAATCACGAAGAAGTTGAAA 726
QY 241 IleCysGluAsp 244
Db 727 ATATGTGAAGAG 738

RESULT 8
AAS34889
ID AAS34889 standard; cDNA; 3707 BP.
XX
AC AAS34889;
XX
DT 04-DEC-2001 (first entry)
XX
DE cDNA encoding novel human neoplastic disease associated polypeptide #123.
KW Human; neoplastic disease associated polypeptide; cancer; gene therapy;
KW hyperproliferative disorder; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
XX neuroprotective; cytostatic; anti inflammatory; vasotropic; ss.
OS Homo sapiens.
XX
PN WO200155163-A1.
XX
PD 02-AUG-2001.
PF 17-JAN-2001; 2001WO-US01359.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465558/50.
DR P-PSDB; AAU21690.
XX
PT Novel polypeptides and polynucleotides useful as diagnostic reagents to
PT diagnose diseases or disorders associated with aberrant expression or
PT activity of polypeptides, and for treating cancers, rheumatoid
PT arthritis -
XX
PS Claim 4; SEQ ID No 133; 687pp; English.
XX
CC The present invention relates to the isolation of novel human neoplastic
CC disease associated polypeptides (AAU21568-AAU21851), and cDNA and DNA
CC sequences encoding for these polypeptides. The sequences of the
CC invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of disorders involving neoplastic disease such as
CC hyperproliferative disorders (e.g. leukaemia, bone cancer, bladder
CC cancer, brain stem glioma, adult liver cancer, childhood cerebellar
CC astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may
CC also be useful for treating other disorders such as neural disorders,
CC immune system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC and renal disorders. The polynucleotide sequences of the invention are
CC also useful in gene therapy. AAS34767-AAS35050 represent cDNA sequences
CC encoding for the novel human neoplastic disease associated polypeptides
CC of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 3707 BP; 936 A; 963 C; 975 G; 833 T; 0 other;
Alignment Scores: 2.48e-101 Length: 3707
Pred. No.: 1278.00 Matches: 240
Score: 100.00% Conservative: 1
Percent Similarity: 99.59% Mismatches: 0
Best Local Similarity: 96.89% Indels: 0
Query Match: 22 Gaps: 0
DB:

US-09-987-755-2 (1-249) x AAS34889 (1-3707)

QY 4 ValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProProTrpPro 23
Db 20 GTGAGTCTCGGGCTCGGCGACTTGGTGTGGGGAAACTCGGCCGATATCTCTTGGCCA 79
QY 24 GlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLysCysPhe 43
Db 80 GGAAGATTGTTAATCCACAAAGGACTTGAAGAAACCTCGCGAAAGAAATGCTTCTTT 139
QY 44 ValLysPheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysProTyr 63
Db 140 GTGAAATTTTGGACAGAAGATCATGCTGGATCAAAAGTGAACAGCTGAAGCCATAT 199
QY 64 HisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnAlaVal 83
Db 200 CATGCTCATAAAGAGGAATGATATAAAATTAACAAGGGTAAACGATTCCAGCAAGCGTA 259
QY 84 AspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSerSerHisAsn 103
Db 260 GATGCTGTCAAGAGTTCTCTCAGGAGAGCCAAAGGGAAGACACGACGTCATCCCAAT 319
QY 104 SerSerAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsnSerGly 123
Db 320 TCTTCTGATGACAAGAAATCGACGTAATTCAGTGAGGAGAGAAGTAGGCCAAACTCAGGT 379
QY 124 AspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGluGly 143
Db 380 GATGAGAAGCGCAAACTTAGCTGTCTGAAGGGAAGGTGAAGAAGAACATGGGAGAAGGA 439
QY 144 LysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArgAla 163
Db 440 AAGAAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAAAGAGCC 499
QY 164 GlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAspLeuThrIle 183
Db 500 CAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGGATGAGAAGGATCTCACCATC 559
QY 184 ProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLysTrpGln 203
Db 560 CCGAGTCTAGTACCGTGAAGGGATGATGCCCGGACCGATGGCCGCTTTAAATGGCAG 619
QY 204 ProThrAlaSerGluProValLysAspAlaAspProHisPheHisPheLeuLeuSer 223
Db 620 CCAACCGCAAGCGAGCCTGTTAAAGATGCAGATCCTCATTTCCATCTTCTGCTAAGC 679
QY 224 GlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLysIleCysGlu 243
Db 680 CAAACAGAGAGCCAGCTGTCTGTACCAGCAATCACGAAGAAGTTGAAATATGTGAA 739
QY 244 Asp 244
Db 740 GAG 742

RESULT 9
AAS33224
ID AAS33224 standard; cDNA; 810 BP.
XX AC AAS33224;
XX DT 04-DEC-2001 (first entry)
XX

DE DNA encoding human secreted protein, Seq ID No 183.
XX Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
KW rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW cytostatic; Alzheimer's disease; parkinson's disease; human; cancer;
KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW Gaucher's disease; neurological disease; cerebrovascular disorder;
XX thrombosis; wound healing; ss.
OS Homo sapiens.
XX WO200155326-A2.
PN 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01347.
PF 31-JAN-2000; 2000US-0179065.
PR (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-451931/48.
DR P-PSDB; AAU20515.
XX New nucleic acids and polypeptides, useful for diagnosing, preventing
PT or treating medical conditions -
PT Claim 1; SEQ ID No 183; 753pp; English.

XX The invention relates to novel isolated nucleic acid molecules (I)
CC encoding human secreted proteins (II). (I) and (II) are used to prevent,
CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits, in
CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
CC the prevention, treatment and diagnosis of diseases associated with
CC inappropriate expression of secreted proteins. (I) and complementary
CC sequences may also be used as DNA probes in diagnostic assays (e.g.
CC polymerase chain reactions (PCR)) to detect and quantitate the presence
CC of similar nucleic acid sequences in samples, and so which patients may
CC be in need of restorative therapy. (II) may also be used as antigens in
CC the production of antibodies and in assays to identify modulators
CC (agonists and antagonists) of the expression and activity of the secreted
CC proteins. The anti-(II) antibodies and antagonists may also be used to
CC down regulate expression and activity of (II). The anti-(II) antibodies
CC may also be used as diagnostic agents for detecting the presence of (II)
CC in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
CC disorders include for example: immune/autoimmune diseases (e.g. HIV
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
CC angina and thrombosis), infections caused by bacteria, viruses and
CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),
CC agonists, antagonists and antibodies can also be used to promote wound
CC healing, maintain organs before transplantation, and support cell culture
CC of primary tissues. AAS33043-AAS33486 represent human secreted protein
CC coding sequences, PCR primers, and related sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification but was obtained in electronic format directly from WIPO
CC at: ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 810 BP; 259 A; 178 C; 218 G; 152 T; 3 other;
Alignment Scores:
Pred. No.: 7.99e-101 Length: 810
Score: 1263.00 Matches: 238
Percent Similarity: 99.17% Conservative: 1
Best Local Similarity: 98.76% Mismatches: 2
Query Match: 95.75% Indels: 0


```
DB:                22                Gaps:                0
US-09-987-755-2 (1-249) x AAS33224 (1-810)

QY      4 ValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProProTrpPro 23
Db      2 GTGAGTCTGCGGCTCGCGGACTTGGTGTGGGGAACCTCGNCGATATCCTCTGGCCA 61

QY      24 GlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLysCysPhePhe 43
Db      62 GGAAGATTGTTAATCCACCAAGGACTTGAAGAACTCGCGGAAGAAATGCTTCTTT 121

QY      44 ValLysPhePheGlyThrGluAspHisAlaTIpIleLysValGluGlnLeuLysProTyr 63
Db      122 GTGAAATTTTITGGACAGAGATCATGCTGGATCAAAGTGAACAGCTGAAGCCATAT 181

QY      64 HisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGlnAlaVal 83
Db      182 CATGCTCATAAAGAGGAATGATAAAATTAAACAGGTAAACGATTCCAGCAAGCGGTA 241

QY      84 AspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSerSerHisAsn 103
Db      242 GATGTGTGCAAGAGTTCCTCAGGAGAGCCAAAGGGAAGACCAGAGTCATCCACAAT 301

QY      104 SerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsnSerGly 123
Db      302 TCITCTGATGACAAAGATCGACGTAATCCAGTGAGGAGAGAGTAGGCCAACTCAGGT 361

QY      124 AspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGluGly 143
Db      362 GATGAGAAGCGCAACTTAGCCTGTCTGAAGGGAAGGTGAAGAACAATGGGAGAAGGA 421

QY      144 LysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArgAla 163
Db      422 AAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTGAAAAGAGCC 481

QY      164 GlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAspLeuThrIle 183
Db      482 CAAGAGCAAAGTYCCGGAAGCGGGGTGGCCCCCAAGGATGAGAAGGATCTCACCATC 541

QY      184 ProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLysTrpGln 203
Db      542 CCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGGTTTAAATGGCAG 601

QY      204 ProThrAlaSerGluProValLysAspAlaAspProHisPheHisPheLeuLeuSer 223
Db      602 CCAACCGCAAGCGAGCCTGTTAAAGATGCAGATCCTCATTTCCATCATTTCTGCTAAGC 661

QY      224 GlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLysIleCysGlu 243
Db      662 CAAACAGAGAAGCCAGCTGTCTGTACCAGGCAATCACGAAGAGTTGAAAATATGTGAA 721

QY      244 Asp 244
Db      :::
Db      722 GAG 724

RESULT 10
AAS35012
ID      AAS35012 standard; cDNA; 810 BP.
XX
AC      AAS35012;
XX
DT      04-DEC-2001 (first entry)
DE      cDNA encoding novel human neoplastic disease associated polypeptide #246.
XX
KW      Human; neoplastic disease associated polypeptide; cancer; gene therapy;
KW      hyperproliferative disorder; neural disorder; immune system disorder;
KW      muscular disorder; reproductive disorder; gastrointestinal disorder;
KW      pulmonary disorder; cardiovascular disorder; renal disorder;
KW      neuroprotective; cytostatic; anti inflammatory; vasotropic; ss.
XX
OS      Homo sapiens.
```

```
XX      WO200155163-A1.
PN
XX      02-AUG-2001.
PD
XX      17-JAN-2001; 2001WO-US01358.
PF
XX      31-JAN-2000; 2000US-0179065.
XX      04-FEB-2000; 2000US-0180628.
XX      24-FEB-2000; 2000US-0184664.
XX      02-MAR-2000; 2000US-0186350.
XX      16-MAR-2000; 2000US-0189874.
XX      17-MAR-2000; 2000US-0190076.
XX      18-APR-2000; 2000US-0198123.
XX      19-MAY-2000; 2000US-0205515.
XX      07-JUN-2000; 2000US-0209467.
XX      28-JUN-2000; 2000US-0214886.
XX      30-JUN-2000; 2000US-0215135.
XX      07-JUL-2000; 2000US-0216647.
XX      07-JUL-2000; 2000US-0216880.
XX      11-JUL-2000; 2000US-0217487.
XX      11-JUL-2000; 2000US-0217496.
XX      14-JUL-2000; 2000US-0218290.
XX      26-JUL-2000; 2000US-0220963.
XX      26-JUL-2000; 2000US-0220964.
XX      14-AUG-2000; 2000US-0224518.
XX      14-AUG-2000; 2000US-0224519.
XX      14-AUG-2000; 2000US-0225213.
XX      14-AUG-2000; 2000US-0225214.
XX      14-AUG-2000; 2000US-0225266.
XX      14-AUG-2000; 2000US-0225267.
XX      14-AUG-2000; 2000US-0225268.
XX      14-AUG-2000; 2000US-0225270.
XX      14-AUG-2000; 2000US-0225447.
XX      14-AUG-2000; 2000US-0225757.
XX      14-AUG-2000; 2000US-0225758.
XX      14-AUG-2000; 2000US-0225759.
XX      18-AUG-2000; 2000US-0226279.
XX      22-AUG-2000; 2000US-0226681.
XX      22-AUG-2000; 2000US-0226688.
XX      22-AUG-2000; 2000US-0227182.
XX      23-AUG-2000; 2000US-0227009.
XX      30-AUG-2000; 2000US-0228924.
XX      01-SEP-2000; 2000US-0229287.
XX      01-SEP-2000; 2000US-0229343.
XX      01-SEP-2000; 2000US-0229344.
XX      01-SEP-2000; 2000US-0229345.
XX      05-SEP-2000; 2000US-0229509.
XX      05-SEP-2000; 2000US-0229513.
XX      06-SEP-2000; 2000US-0230437.
XX      06-SEP-2000; 2000US-0230438.
XX      08-SEP-2000; 2000US-0231242.
XX      08-SEP-2000; 2000US-0231243.
XX      08-SEP-2000; 2000US-0231244.
XX      08-SEP-2000; 2000US-0231413.
XX      08-SEP-2000; 2000US-0231414.
XX      08-SEP-2000; 2000US-0232080.
XX      08-SEP-2000; 2000US-0232081.
XX      12-SEP-2000; 2000US-0231968.
XX      14-SEP-2000; 2000US-0232397.
XX      14-SEP-2000; 2000US-0232398.
XX      14-SEP-2000; 2000US-0232399.
XX      14-SEP-2000; 2000US-0232400.
XX      14-SEP-2000; 2000US-0232401.
XX      14-SEP-2000; 2000US-0233063.
XX      14-SEP-2000; 2000US-0233064.
XX      14-SEP-2000; 2000US-0233065.
XX      21-SEP-2000; 2000US-0234223.
XX      21-SEP-2000; 2000US-0234274.
XX      25-SEP-2000; 2000US-0234997.
XX      25-SEP-2000; 2000US-0234998.
XX      26-SEP-2000; 2000US-0235484.
XX      27-SEP-2000; 2000US-0235834.
PR
```


PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX

PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465558/50.
DR P-PSDB; AAU21813.
XX
PT Novel polypeptides and polynucleotides useful as diagnostic reagents to
PT diagnose diseases or disorders associated with aberrant expression or
PT activity of polypeptides, and for treating cancers, rheumatoid
PT arthritis
XX
PS Claim 4; SEQ ID No 256; 687pp; English.
XX
CC The present invention relates to the isolation of novel human neoplastic
CC disease associated polypeptides (AAU21568-AAU21851), and cDNA and DNA
CC sequences encoding for these polypeptides. The sequences of the
CC invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of disorders involving neoplastic disease such as
CC hyperproliferative disorders (e.g. leukaemia, bone cancer, bladder
CC cancer, brain stem glioma, adult liver cancer, childhood cerebellar
CC astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may
CC also be useful for treating other disorders such as neural disorders,
CC immune system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC and renal disorders. The polynucleotide sequences of the invention are
CC also useful in gene therapy. AAS34767-AAS35050 represent cDNA sequences
CC encoding for the novel human neoplastic disease associated polypeptides
CC of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 810 BP; 259 A; 178 C; 218 G; 152 T; 3 other;

Alignment Scores:

Pred. No.: 7.99e-101 Length: 810
Score: 1263.00 Matches: 238
Percent Similarity: 99.17% Conservative: 1
Best Local Similarity: 98.76% Mismatches: 2
Query Match: 95.75% Indels: 0
DB: 22 Gaps: 0

US-09-987-755-2 (1-249) x AAS35012 (1-810)

Qy 4 ValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProTrpPro 23
Db 2 GTGAGTCTGCGNGCTCGGCGACTTGGTGTGGGGGAACTCGGCGATATCTCTTGCCCA 61
Qy 24 GlyLysIleValAsnProLysAspLeuLysLysProArgGlyLysLysCysPhe 43
Db 62 GGAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAATGCTTCTT 121
Qy 44 ValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysProTyr 63
Db 122 GTGAAATTTTGGACAGAGATCATGCTGGATCAAAGTGAACAGCTGAAGCCATAT 181
Qy 64 HisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnAlaVal 83
Db 182 CATGCTCATAAGAGGAATGATAAAATTAACAAGGGTAAACGATTCCAGCAAGCGTA 241
Qy 84 AspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSerSerHisAsn 103
Db 242 GATGCTGTGGAAGATTCTCTCAGGAGAGCCAAAGGAAAGACCAGACGTATCCCAAT 301
Qy 104 SerSerAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsnSerGly 123
Db 302 TCTTCTGATGACAAGATCGACGTAATTCAGTGAGGAGAGAGTAGGCCAACTCAGGT 361
Qy 124 AspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGluGly 143
Db 362 GATGAGAGCGCAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAACATGGGAGAGGA 421
Qy 144 LysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArgAla 163

Db 422 AAGNAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAAAAAGAGCC 481

Qy 164 GlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAspLeuThrIle 183

Db 482 CAAGAGCAAAAGTYCCCGGAAGCGGGTCGGCCCCCAAGAGATGAGAAGGATCTCACCATC 541

Qy 184 ProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLysTrpGln 203

Db 542 CCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCGTTTAAATGGCAG 601

Qy 204 ProThrAlaSerGluProValLysAspAlaAspProHisPheHisHisPheLeuLeuSer 223

Db 602 CCAACCGCAAGCGAGCTGTTAAAGATGCAGATCCTCATTTCCATCATTTCTGTCTAAGC 661

Qy 224 GlnThrGluLysProAlaValCysTyrrGlnAlaIleThrLysLysLeuLysIleCysGlu 243

Db 662 CAAACAGAGAAGCCAGCTGTCTGTTACCAGGCAATCACGAAGAAGTTGAAAAATATGTGA 721

Qy 244 Asp 244

Db 722 GAG 724

RESULT 11

AAV87516/c

ID AAV87516 standard; cDNA; 658 BP.

XX

AC AAV87516;

XX

DT 27-APR-1999 (first entry)

XX

DE EST clone BV27.

XX

KW Expressed sequence tag; secreted protein; haematopoiesis regulator;

KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;

KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;

KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

XX

OS Homo sapiens.

XX

PN WO9845435-A2.

XX

PD 15-OCT-1998.

XX

PF 10-APR-1998; 98WO-US06954.

XX

PR 10-APR-1997; 97US-0835913.

XX

PA (GEMV) GENETICS INST INC.

XX

PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

PI Racie LA, Spaulding V, Treacy M;

XX

DR WPI; 1999-070076/06.

XX

PT New polynucleotides encoding human secreted proteins - derived from

PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,

PT ovary, pituitary, retina and colon cDNA libraries

XX

PS Claim 1; Page 589; 633pp; English.

XX

CC This sequence represents an expressed sequence tag (EST), and is a

CC polynucleotide of the invention. The polynucleotides of the invention are

CC all secreted EST sequences isolated from a variety of human tissue

CC sources. The EST sequences and proteins encoded by them are predicted to

CC have useful biological activities which would make them suitable for

CC treating, preventing or ameliorating medical conditions in humans and

CC animals, although no supporting data is given. Suggested activities

CC include nutritional activity, immune stimulating or suppressing activity,

CC haematopoiesis regulating activity, tissue growth activity,

CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory

CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition

CC activity. The EST sequences are also stated to be useful for gene

CC therapy.

XX

SQ Sequence 658 BP; 148 A; 176 C; 129 G; 205 T; 0 other;

Alignment Scores:

Pred. No.: 1.46e-42 Length: 658

Score: 591.50 Matches: 118

Percent Similarity: 97.54% Conservative: 1

Best Local Similarity: 96.72% Mismatches: 2

Query Match: 44.84% Indels: 1

DB: 20 Gaps: 1

US-09-987-755-2 (1-249) x AAV87516 (1-658)

Qy 91 ArgArgAlaLysGlyLysAspGlnThrSerSerHisAsnSerSerAspLysAsnArg 110

Db 375 AGGAGAGCCAAAGGGAAGACAGACGTCATCCCAATTTCTGTATGACAAATCGA 316

Qy 111 ArgAsnSerSerGluGluArgSerArgProAsnSerGlyAspGluLysArgLysLeuSer 130

Db 315 CGTAATTCAGTGAGGAGAGAAGTAGGCCAAACTCAGGTGATGAGNAGCGCAAACTTAGC 256

Qy 131 LeuSerGluGlyLysValLysLysAsnMetGlyGluGlyLysLysArgValSerSerGly 150

Db 255 CTGTCTGAAGGGAAGGTGAAGAAGAACATGGGAGAAGGAAAGAGAGGGTGTCTCAGGC 196

Qy 151 SerSerGluArgGlySerLysSerProLeuLysArgAlaGlnGluGlnSerProArgLys 170

Db 195 TCTTCAGAGAGAGGCTCCAAATCCCCTCTGAAAAGAGCCCAAGAGCAAAAGTCCCGGAAG 136

Qy 171 ArgGlyArgProProLysAspGluLysAspLeuThrIleProGluSerSerThrValLys 190

Db 135 CGGGTCCGCCCCCAAGGATGAGAAGGATCTCAGCATCCCGGAGTCTAGTACCGTGAAG 76

Qy 191 GlyMetMetAlaGlyProMetAlaAlaPheLysTrpGlnProThrAla---SerGluPro 209

Db 75 GGGATGATGCGCGGACCGATGGCCCGGTTTAAATGGCAGCCCAACCGCGTCTTAGGCCA 16

Qy 210 ValLys 211

Db 15 TGAAAG 10

RESULT 12

AAK70375

ID AAK70375 standard; DNA; 2650 BP.

XX

AC AAK70375;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25187.

XX

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

KW cytostatic; gene therapy; vaccine; metastasis; ds.

XX

OS Homo sapiens.

XX

PN WO200157182-A2.

XX

PD 09-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US01354.

XX

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and
metastasis -

Disclosure; SEQ ID NO 25187; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
activity, and can be used in gene therapy and vaccine production. (I)
proteins, and polynucleotides may be used in the prevention, diagnosis and
treatment of diseases associated with inappropriate (I) expression. For
example, they may be used to treat disorders associated with decreased
expression by rectifying mutations or deletions in a patient's genome

CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 2650 BP; 754 A; 509 C; 779 G; 608 T; 0 other;

Alignment Scores:

Pred. No.: 3.15e-34 Length: 2650
Score: 504.00 Matches: 119
Percent Similarity: 39.24% Conservative: 5
Best Local Similarity: 37.66% Mismatches: 8
Query Match: 38.21% Indels: 184
DB: 22 Gaps: 1

US-09-987-755-2 (1-249) x AAK70375 (1-2650)

Qy 62 ProTyrHisAlaHisLysGluMetIleLysLeuAsnLysGlyLysArgPheGlnGln 81
Db 2 CCATATCATGCTCATAAAGAGGAATGATAAAATAACCAAGGTAAACGATCCAGCAA 61
Qy 82 AlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAsp----- 97
Db 62 GCGGTAGATGCTGTCGAAGAGTTCCTCAGGAGAGCAAGGAAAGCAAGCAGGTGAGAGAC 121
Qy 97 ----- 97
Db 122 ACAATGTTCACTTTTCTACCAATGCCAGACTGGTGATTTTTCATAGCAACTAACAGATTA 181
Qy 97 ----- 97
Db 182 AATGTCACCCAAACGCAAACTTTGTTCTTCTGAAATGCTACTGAGAAAGCTACTTCAG 241
Qy 97 ----- 97
Db 242 TTCTAAGGTGTGAGCGTTTTCGTTATCATGTTTGAACCTCCATGCAGCTAACCTTCTGGA 301
Qy 97 ----- 97
Db 302 GTAGGTGTTGCTTTTAGGAAATGCTTGGACCCAGGAGAAATATCTCAGCATTTGAGT 361
Qy 97 ----- 97
Db 362 CTAGCACAGGCCAATTTCTTCCTTTCATAGCACACAGATTAGGAGCATAGCCTCTGAAATA 421
Qy 97 ----- 97
Db 422 CATATACTGCCTAGATTAAATCCTTTCTTCTAGTTTGGGACATAGGAAAGTGACTTT 481
Qy 97 ----- 97
Db 482 TGATATATATATTTTAAATCTGCTTCTCCCAATGCTGCGGACGACAGCAATCTATT 541
Qy 97 ----- 97
Db 542 TCTAGGGTCATTGTGAGGAGACCACACAGAGTGGCTGTAGCAGCATGCGCACTGTAAA 601
Qy 97 ----- 97
Db 602 TGATGGTGGCGATGGCGGCAGCAGCTGTGGTGGTAGTTATCATTTATTATAAATTT 661
Qy 98 GlnThrSerSerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArg 117
Db 662 CAGACGTATCCCAACAATTTCTCTGATGACAAGATCGACGTAATCCAGTGGAGAGAGA 721
Qy 118 SerArgProAsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysVallys 137

Db 722 AGTAGGCCAAACTCAGGTGATGATGAGAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAG 781
Qy 138 LysAsnMetGlyGlyGlyLysLysArgValSerSerGlySerSerGluArgGlySerLys 157
Db 782 AAGAACAATGGAGAGAGGAAGAAAGAGGGTCTCTTCAGGCTCTTCAGAGAGAGGCTCCAAA 841
Qy 158 SerProLeuLysArgAlaGlnGlnSerProArgLysArgGlyArgProProLysAsp 177
Db 842 TCCCTCTCTGAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCAAGGAT 901
Qy 178 GluLysAspLeuThrIleProGluSerSerThrValLysGlyMetMet 193
Db 902 GAGAAGGTTTGTCTTCTCTCTGCAGCTTCCGGATTAGTGATGTCCTT 949
RESULT 13
AAK70376
ID AAK70376 standard; DNA; 2650 BP.
XX
AC AAK70376;
XX
DT 06-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25188.
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytosstatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.

PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-483426/52.
DR
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 25188; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 2650 BP; 754 A; 509 C; 779 G; 608 T; 0 other;

Alignment Scores:
Pred. No.: 3.15e-34 Length: 2650
Score: 504.00 Matches: 119
Percent Similarity: 39.24% Conservative: 5
Best Local Similarity: 37.66% Mismatches: 8
Query Match: 38.21% Indels: 184
DB: 22 Gaps: 1

US-09-987-755-2 (1-249) x AAK70376 (1-2650)

Qy 62 ProTyrHisAlaHisLysGluMetIleLysIleAsnLysGlyLysArgPheGlnGln 81
Db 2 CCATATCATGCTCATAAGAGGAATGATAAAATTAACAAGGGTAAACGATTCAGCAA 61
Qy 82 AlaValAspAlaValGluPheLeuArgAlaLysGlyLysAsp----- 97

```
Db 62 GCGGTAGATGCTGTGGAAGAGTTCTCTCAGGAGAGCCAAAGGAAAGACCCAGGTGAGAGAC 121
QY 97 -----
Db 122 ACAATGTTCACTTTCTACCAATGCCAGACTGGTGATTTTGCATAGCAACTAACAGATTA 181
QY 97 -----
Db 182 AATGTCACCCAAACGCAAAACCTTTGTTCTTCTGAAATGCTACTGGAGAAAGCTACTTCAG 241
QY 97 -----
Db 242 TTCTAAGGTGTCAGCGTTTTCCTTATCAGTTTGAAACTCCATGCAGCTAACCTTTCTGGA 301
QY 97 -----
Db 302 GTAGTTGTTGCTTTTAGGAAAATTGCTTGGACCCAGGAGAAATATCTCAGCATTTGAGT 361
QY 97 -----
Db 362 CTAGCACAGGCCAATTCTTCTTCATAGCACACAGATTAGGAGCATAGCCTCTGAAATA 421
QY 97 -----
Db 422 CATATACTGCCTAGATTAAATCCTTTCTTCTAGTTTGGACATAGGAAAGTGACTTT 481
QY 97 -----
Db 482 TGATATATATATTTTAAATCTGCTTCTCCCAATGCTCTGGACGACAGAGATCTATT 541
QY 97 -----
Db 542 TCTAGGTCATTGTGAGGAGAGACCACACAGAGTGCTGTAGCAGCATGGCCTGTAAA 601
QY 97 -----
Db 602 TGATGTTGGCGATGGCGGCAGCAGCTGTGGTGTAGTTATCTATTATTAAACTTT 661
QY 98 GlnThrSerSerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArg 117
Db 662 CAGACGTCATCCCAACAATTCTTCTGATGACAAGATCGACGTAATCCAGTGAGGAGAGA 721
QY 118 SerArgProAsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLys 137
Db 722 AGTAGGCCAAACTCAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAG 781
QY 138 LysAsnMetGlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLys 157
Db 782 AAGAACATGGGAGAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAA 841
QY 158 SerProLeuLysArgAlaGlnGlnSerProArgLysArgGlyArgProProLysAsp 177
Db 842 TCCCTCTGAAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGGAT 901
QY 178 GluLysAspLeuThrIleProGluSerSerThrValLysGlyMetMet 193
Db 902 GAGAAAGTTTGTCTCTCTCTGACGCTTCCGGGATTAGTGATGTCCTT 949
```

RESULT 14

AAV86389/c
ID AAV86389 standard; cDNA; 307 BP.

XX AAV86389;

XX AC AAV86389;

XX 27-APR-1999 (first entry)

XX EST clone AQ73.

XX Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

```
XX OS Homo sapiens.
XX PN WO9845435-A2.
XX PD 15-OCT-1998.
XX PF 10-APR-1998; 98WO-US06954.
XX PR 10-APR-1997; 97US-0835913.
XX PA (GEMY ) GENETICS INST INC.
XX PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX WPI; 1999-070076/06.
XX New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
XX Claim 1; Page 218; 633pp; English.
XX This sequence represents an expressed sequence tag (EST), and is a
CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
XX therapy.
SQ Sequence 307 BP; 67 A; 91 C; 62 G; 85 T; 2 other;
Alignment Scores:
Pred. No.: 3.87e-30 Length: 307
Score: 444.00 Matches: 80
Percent Similarity: 97.59% Conservative: 1
Best Local Similarity: 96.39% Mismatches: 2
Query Match: 33.66% Indels: 0
DB: 20 Gaps: 0
US-09-987-755-2 (1-249) x AAV86389 (1-307)
QY 1 MetAlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
Db 271 ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCCGATATCCT 212
QY 21 ProTrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLys 40
Db 211 CCTTGGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAA 152
QY 41 CysPhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu 60
Db 151 TGCTTCTTTGTGAAATTTTGTGAACAGAGATCATGCCTGGATCAAAGTGAACAGCTG 92
QY 61 LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln 80
Db 91 AAGCCATATCATGCTCATAAAGAGGAGANATGATAAAATTAACNAGGGTAAACGATTCCAG 32
QY 81 GlnAlaVal 83
Db 31 CAAGCACTC 23
RESULT 15
AAC21457
```


ID AAC21457 standard; cDNA; 445 BP.
XX AAC21457;
AC
XX
XX 06-OCT-2000 (first entry)
DT
XX Human secreted protein 5' EST, SEQ ID NO: 25532.
DE
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
OS
XX EP1033401-A2.
PN
XX 06-SEP-2000.
PD
XX 21-FEB-2000; 2000EP-0200610.
PF
XX 26-FEB-1999; 99US-0122487.
PR
XX (GEST) GENSET.
FA
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
DR
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 25532; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 445 BP; 123 A; 90 C; 129 G; 103 T; 0 other;

Alignment Scores:
Pred. No.: 5.43e-21 Length: 445
Score: 341.00 Matches: 71
Percent Similarity: 51.80% Conservative: 1
Best Local Similarity: 51.08% Mismatches: 0
Query Match: 25.85% Indels: 67
DB: 21 Gaps: 1

US-09-987-755-2 (1-249) x AAC21457 (1-445)

Qy 1 MetAlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
Db 18 ATGGCGGCTGTGAGTCTGCGGCTCGGCGACTGGTGTGGGGAAACCTCGGCGGATATCCT 77
Qy 21 ProTrpProGlyLysIleValAsnProProLysAspLeuLysProArgGlyLysLys 40
Db 78 CCTTGGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAA 137
Qy 41 CysPhePheValLysPhePheGlyThrGluAsp-His----- 52
Db 138 TGCTTCTTTGTGAAATTTTGGACACAGAGATCAATAGACCAGCCCCATTCTCTATC 197
Qy 52 ----- 52

Db 198 ACAAGCCAGCTTAGCAGCCTGTGGTGGCAGTATTTCAAGGAGATCGCTGGCCTGGAGCTG 257
Qy 52 ----- 52
Db 258 TGGAGCCAGAGCTTCTGGAAGAGGGAATGTTCTTAGAACTCCATGGATTAAAAAGAAGA 317
Qy 53 -----AlaTr 54
Db 318 TTTGGCCGGGCATGGTGCCTCAGGCCTGCAATCCCAGTACTTTGGGAGGCCCAAGTGCCTG 377
Qy 54 pIleLysValGluGlnLeuLysProTyrHisAlaHisLysGluGluMetIleLys 72
Db 378 GATCAAAGTGGACACAGCTGAAGCCATATCATGCTCATAAAGAGGAAATGATATAAA 432

Search completed: December 8, 2003, 11:40:26
Job time : 285 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n_model

Run on: December 8, 2003, 11:35:47 ; Search time 2151 Seconds
(without alignments)
2813.489 Million cell updates/sec

Title: US-09-987-755-2
Perfect score: 1319
Sequence: 1 MAAVSLRLGDLVWGKLGryp.....VCYQAITKKLKICEDLLLP 249

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09987755/runat_08122003_103436_18319/app_query_fasta_1.391
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09987755 @CGN_1_1_2810 @runat_08122003_103436_18319 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_estfun:*
15: em_estom:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1291	97.9	1201	13	BX331741 BX331741
2	1287	97.6	3741	11	BC032855 Homo sapi
3	1283	97.3	1039	12	BM808810 AGENCOURT
4	1282	97.2	870	13	BQ720854 AGENCOURT
5	1282	97.2	884	13	BU158924 AGENCOURT
6	1257	95.3	1003	9	AL536040 AL536040
7	1244.5	94.4	1005	12	BM808969 AGENCOURT
8	1236.5	93.7	993	13	BU520128 AGENCOURT
9	1232.5	93.4	890	13	BQ918182 AGENCOURT
10	1228.5	93.1	899	13	BQ892009 AGENCOURT
11	1225.5	92.9	3239	11	AK014456 Mus muscu
12	1222	92.6	755	12	BI551885 603192793
13	1214.5	92.1	813	12	BI691175 603314566
14	1214.5	92.1	929	13	BQ934274 AGENCOURT
15	1211.5	91.8	3250	11	AK040841 Mus muscu
16	1207.5	91.5	829	13	BU593961 AGENCOURT
17	1202.5	91.2	705	13	BU708303 UI-M-FCO-
18	1164	88.2	778	12	BI600148 603246545
19	1160.5	88.0	773	12	BI103792 602887976
20	1159	87.9	816	12	BI459800 603200852
21	1155.5	87.6	704	13	BQ768940 UI-M-FCO-
22	1151.5	87.3	801	14	CA512495 UI-R-FJO-
23	1144.5	86.8	685	10	BB617747 BB617747
24	1144	86.7	816	13	BU373180 603589670
25	1144	86.7	1138	12	BM906992 AGENCOURT
26	1130.5	85.7	682	14	BY735423 BY735423
27	1127	85.4	998	10	BB611224 BB611224
28	1111.5	84.3	823	12	BI150279 602915496
29	1106.5	83.9	815	12	BI767890 603061025
30	1093.5	82.9	732	12	BI218947 602938548
31	1077.5	81.7	671	12	BI732831 603353939
32	1069.5	81.1	659	12	BM947061 UI-M-EHOP
33	1064.5	80.7	660	14	BY753847 BY753847
34	1052.5	79.8	646	14	BY745429 BY745429
35	1043.5	79.1	786	12	BG976825 602845727
36	1035	78.5	775	14	CB601170 AGENCOURT
37	1026	77.8	861	12	BI689620 603316113
38	1018	77.2	710	9	AJ453917 AJ453917
39	1016.5	77.1	595	14	CB584042 AMGNNUC:N
40	1002	76.0	702	9	AJ455240 AJ455240
41	989	75.0	671	9	AJ456583 AJ456583
42	985	74.7	821	12	BG870714 602792108
43	977	74.1	908	10	BG294400 602391505
44	959	72.7	642	14	BY745902 BY745902
45	953	72.3	868	13	BU358607 603477645

ALIGNMENTS

RESULT 1
BX331741
LOCUS
DEFINITION BX331741 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
CDNA clone CS0DC001YL09 5-PRIME, mRNA sequence.
ACCESSION BX331741
VERSION BX331741.1 GI:30335078
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)

AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7982.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DC001CF05QP1&cluster=7982.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InVtroGen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DC001CF05QP1.
FEATURES Location/Qualifiers
source 1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC001YL09"
/tissue type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 349 a 261 c 332 g 240 t 19 others
ORIGIN
Alignment Scores:
Pred. No.: 4.18e-96 Length: 1201
Score: 1291.00 Matches: 243
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 97.88% Indels: 0
DB: 13 Gaps: 0
US-09-987-755-2 (1-249) x BX331741 (1-1201)
QY 1 MetAlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
Db 75 ATGCGGGCTGTAGTCTCGGCTCGCGACTTGGTGTGGGGAAACTCGGCCGATATCCT 134
QY 21 ProTrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLys 40
Db 135 CCTGGCCAGGAAGATTGTHAATCCACCAAGGACTTGAAGAAACCTCGCGGAAGAAA 194
QY 41 CysPhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu 60
Db 195 TGCTCTTTGTGAAATTTTGTGAAACAGAGATCATGCTGGATCAAAAGTGAACAGCTG 254
QY 61 LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln 80
Db 255 AAGCCATATCATGCTCATAAGAGGAAATGATAAAAATTAACAAGGGTAAACGATTCCAG 314
QY 81 GlnAlaValAspAlaValGluPheLeuArgArgAlaLysGlyLysAspGlnThrSer 100
Db 315 CAACGGGTAGATGTGTCGAAGAGTTCCTCAGGAGAGCCAAAGGAAAGACCAGACGTCA 374
QY 101 SerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro 120
Db 375 TCCACAAATTCTTCTGATGACAAAGAATCGACGCTAATTCAGTGAGGAGAGAGTAGGCCA 434
QY 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysAsnMet 140
Db 435 AACTCAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAACAATG 494
QY 141 GlyGluGlyLysLysArgValSerSerGlySerGluArgGlySerLysSerProLeu 160
Db 495 GGAGAAGGAAGAGAGGGGTGCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG 554
QY 161 LysArgAlaGlnGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp 180

Db 555 AAAAGAGCCCAAGAGCAAAGTCCCCGGAAGCGGGTCTGGCCCCCAAGGATGAGAAGGAT 614
QY 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
Db 615 CTCACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCTTT 674
QY 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
Db 675 AAATGGCAGCCCAACCGCAAGCGAGCTGTAAAGATGCAGATCCTCATTTCCATCATTC 734
QY 221 LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys 240
Db 735 CTGCTAAGCCAAACAGAGAGACCGAGCTGTCTGTATTACCAGCAATCACGAAGAAGTTGAA 794
QY 241 IleCysGluAsp 244
Db 795 ATATGTGAAGAG 806
RESULT 2
BC032855
LOCUS BC032855 3741 bp mRNA linear HTC 04-MAR-2003
DEFINITION Homo sapiens, Similar to RIKEN cDNA 3930401K13 gene, clone
IMAGE:5265561, mRNA.
ACCESSION BC032855
VERSION BC032855.1 GI:23958784
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3741)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAX Plate: 47 Row: j Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein
This clone has the following problem: retained intron.
FEATURES Location/Qualifiers
source 1..3741
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5265561"
/tissue type="Testis"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"
BASE COUNT 940 a 959 c 993 g 849 t
ORIGIN
Alignment Scores:
Pred. No.: 2.74e-95 Length: 3741

Score:	1287.00	Matches:	242
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	99.18%	Mismatches:	0
Query Match:	97.57%	Indels:	0
DB:	11	Gaps:	0
US-09-987-755-2 (1-249) x BC032855 (1-3741)			
QY	1	MetAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro	20
Db	29	ATGGCGCTGTGAGTCTCGGCTCGGCTGTTGGTGGGGAACTCGGCGGATATCCT	88
QY	21	ProTrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLys	40
Db	89	CCTTGGCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAACTCGCGGAAAGGAA	148
QY	41	CysPhePheValLysPheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu	60
Db	149	TGCTTCTTTGTGAAATTTTGGAAACAGAGATCATGCTGGATCAAAGTGAACAGCTG	208
QY	61	LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln	80
Db	209	AAGCCATATCATGCTCATAAAGAGGAATGATAAAATTAACAAGGGTAAACGATTCCAG	268
QY	81	GlnAlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSer	100
Db	269	CAAGCGGTAGATGCTGTGGAAGATTCTCTCAGGAGAGCCAAAGGAAAGACCAGCTCA	328
QY	101	SerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro	120
Db	329	TCCCACATCTCTTGATGACAAAGATCGACGTAATTCAGTGAGGAGAGTAGGCCA	388
QY	121	AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet	140
Db	389	AACTCAGTGATGAGAAAGCGCAACTTAGCCTGTCTGAGGGAGGTGAAGAAGCATG	448
QY	141	GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu	160
Db	449	GGAGAGGAAAGAGAGGGTGTCTTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCCTG	508
QY	161	LysArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp	180
Db	509	AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGGGGTGCGGCCCCCAAGGATGAGAAGAT	568
QY	181	LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe	200
Db	569	CTCACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGCGCGGACCGATGGCCGCTT	628
QY	201	LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe	220
Db	629	AAATGGCAGCAACCGCAAGGAGCCTGTTAAAGATGCAATCTCATTTCCATCATTC	688
QY	221	LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys	240
Db	689	CTGCTAAGCCAAACAGAGAGCCAGCTGTCTGTATTACCAGCAATCACGAAGAAGTTGAA	748
QY	241	IleCysGluAsp	244
Db	749	ATATGTGAAGAG	760
RESULT 3	BM808810	1039 bp	linear
LOCUS	AGENCOURT_6617726	NIH_MGC_124	Homo sapiens
DEFINITION	5', mRNA sequence.		EST 05-MAR-2002
ACCESSION	BM808810		EST 5734324
VERSION	BM808810.1	GI:19125633	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 1039)		

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Invitrogen
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM12739 row: h column: 05
High quality sequence stop: 715.
Location/Qualifiers
1. .1039
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5734324"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_124"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: EcoRV (destroyed); Site_2: NotI; RNA source male hippocampus, age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 0.9-4 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 012."

BASE COUNT 307 a 247 c 288 g 197 t

ORIGIN

Alignment Scores:
Pred. No.: 1.64e-95 Length: 1039
Score: 1283.00 Matches: 241
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 99.18% Mismatches: 0
Query Match: 97.27% Indels: 0
DB: 12 Gaps: 0

US-09-987-755-2 (1-249) x BM808810 (1-1039)

QY 2 AlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProPro 21
Db 9 GCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCGGATATCCTCT 68

QY 22 TrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLysCys 41
Db 69 TGGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAATGC 128

QY 42 PhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLys 61
Db 129 TTCTTTGTGAAATTTTGTGAACAGAAAGATCATGCTCGATCAAAGTGAACAGCTGAAG 188

QY 62 ProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGln 81
Db 189 CCATATCATGCTCATAAACAGGAAATGATAAAATTAACAAGGGTAAACGATTCCAGCAA 248

QY 82 AlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSerSer 101
Db 249 GCGGTAGATGCTGTGGAAGAGTTCTTCAGGAGAGCCAAAGGAAAGACAGACGTCATCC 308

QY 102 HisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsn 121
Db 309 CACAATTCTTCTGATGACAAGAATCGACGTAATTCAGTGAGGAGAGAGTAGGCCAAAC 368

QY 122 SerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGly 141
Db 369 TCAGGTGATGAGAGGCAAACTTAGCCTGTCTGAAGGAAAGGTGAAGAAGACATGGGA 428

QY 142 GluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLys 161

Db 429 GAAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTGAAA 488
Qy 162 ArgAlaGlnGluSerProArgLysArgGlyArgProProLysAspGluLysAspLeu 181
Db 489 AGAGCCCAAGAGCAAGTCCCGGAAGCGGGTCCGCCCAAGGATGAGAGGATCTC 548
Qy 182 ThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLys 201
Db 549 ACCATCCCGAGTGTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCTTTAAA 608
Qy 202 TrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPheLeu 221
Db 609 TGGCAGCCCAACCGCAAGCGAGCCCTGTAAAGATGCAGATCCTCATTTCCATCATTCCTG 668
Qy 222 LeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLysIle 241
Db 669 CTAAGCCCAACAGAGAAGCCAGCTGTCTGTACCAGGCAATCACGAAGAAGTTGAAAATA 728
Qy 242 CysGluAsp 244
Db 729 TGTGAAGAG 737

RESULT 4
BQ720854
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BQ720854 870 bp mRNA linear EST 16-JUL-2002
AGENCOURT 8241262 Lupski sympathetic_trunk Homo sapiens cDNA clone
IMAGE:6187142 5', mRNA sequence.
BQ720854
BQ720854.1 GI:21859751
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13580 row: k column: 15
High quality sequence stop: 668.

FEATURES
source
1..870
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6187142"
/sex="male"
/tissue_type="sympathetic trunk"
/dev_stage="adult, 16 yr"
/lab_host="DH10B"
/clone_lib="Lupski sympathetic trunk"
/note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site 2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGCGTCCG-3' and
5'-GACTAGTTCTAGATCGGAGCGGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.9 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine); available through Life
Technologies."
270 a 193 c 245 g 161 t 1 others

BASE COUNT
ORIGIN

Alignment Scores:
Pred. No.: 1.67e-95 Length: 870
Score: 1282.00 Matches: 241
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 97.19% Indels: 0
DB: 13 Gaps: 0

US-09-987-755-2 (1-249) x BQ720854 (1-870)

Qy 3 AlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProProTrp 22
Db 3 GCTGTGAGTCTCGGGCTCGGCGACTTGGTGTGGGGGAAACTCGGCCGATATCCTCTTGG 62
Qy 23 ProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLysCysPhe 42
Db 63 CCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAATGCTTC 122
Qy 43 PheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysPro 62
Db 123 TTTGTGAATTTTGTGAACAGAAAGATCATGCTGGATCAAAAGTGAACAGCTGAAGCCA 182
Qy 63 TyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGlnAla 82
Db 183 TATCATGCTCATAAAGAGGAAATGATAAAATTAACAAGGTAAACGATTCCAGCAAGCG 242
Qy 83 ValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSerSerHis 102
Db 243 GTAGATGCTGTGGAAGAGTTCTTCAGGAGAGCCAAAGGGAAGACCAAGACGTCATCCCAC 302
Qy 103 AsnSerSerAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsnSer 122
Db 303 AATTCTTCTGATGACAAGAAATCGACGTAATTCAGTGAGGAGAGAGTAGGCCAAACTCA 362
Qy 123 GlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGlu 142
Db 363 GGTGATGAGAAGCGCAAACTTAGCCTGTCTGAAGGGAAGTGAAGGAAGAACATGGGAGAA 422
Qy 143 GlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArg 162
Db 423 GGAAGAAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTGAAAAAGA 482
Qy 163 AlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAspLeuThr 182
Db 483 GCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCAAGGATGAGAAGGATCTCACC 542
Qy 183 IleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLysTrp 202
Db 543 ATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCTTTAAATGG 602
Qy 203 GlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPheLeuLeu 222
Db 603 CAGCCCAACCGCAAGCGAGCCTGTAAAGATGCAGATCCTCATTTCCATCTCTGCTA 662
Qy 223 SerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLysIleCys 242
Db 663 AGCCAAACAGAGAAGCCAGCTGTCTGTTACCAGGCAATCACGAAGAAGTTGAAAATATGT 722
Qy 243 GluAsp 244
Db 723 GAAGAG 728

RESULT 5

BUI58924
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BUI58924 884 bp mRNA linear EST 04-SEP-2002
AGENCOURT_7844927 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6044084
5', mRNA sequence.
BUI58924
BUI58924.1 GI:22672834
EST.
Homo sapiens (human)
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 884)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAW13286 row: b column: 21
High quality sequence stop: 776.

Location/Qualifiers
1. .884
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6044084"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 92"
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 274 a 195 c 244 g 169 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 1.69e-95 Length: 884
Score: 1282.00 Matches: 241
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 97.19% Indels: 0
DB: 13 Gaps: 0

US-09-987-755-2 (1-249) x BU158924 (1-884)

QY 3 AlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProProTrp 22
Db GCTGTGAGTCTGGGCTCGGCGACTTGGTGGGGGAACTCGGCCGATATCTCTCTGG 61

QY 23 ProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLysCysPhe 42
Db CCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAACTCGCGGAAAGAAATGCTTC 121

QY 43 PheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysPro 62
Db TTTGTGAAATTTTGTGAACAGAAAGATCATGCTGGATCAAAAGTGAACAGCTGAAGCCA 181

QY 63 TyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnAla 82
Db TATCATGCTCTAAAGAGGAATGATATAAAATAACAAAGGGTAAACGATTCAGCAAGCG 241

QY 83 ValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSerSerHis 102
Db GTAGATGCTGTGCAAGAGTTCTCAGGAGAGCAAGGAAAGGAAAGACCAAGCTATCCAC 301

QY 103 AsnSerSerAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsnSer 122
Db AATTCTTCTGATGACAAAGATCGACGTAAATTCAGTGAGGAGAGAGTAGGCCAACTCA 361

QY 123 GlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGlu 142
Db GGTGATGAGAGCGCAACTTAGCTGTCTGAAGGGAAGGTGAAGAGAACATGGGAGAA 421

QY 143 GlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArg 162

Db 422 GGAAGAAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAAAGA 481

QY 163 AlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGlyLysAspLeuThr 182
Db GCCCAAGAGCAAAGTCCCCGGAAGCGGGTCTGGCCCCCAAGGATGAGAAGGATCTCACC 541

QY 183 IleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLysTrp 202
Db ATCCCGGAGTCTAGTACCGTGAAGGGATGATGGCCGACCGATGGCGCGTTTAAATGG 601

QY 203 GlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPheLeuLeu 222
Db CAGCCAAACCGAAGCGAGCCTGTTAAAGATGCAGATCTCATTTCCATCATTTCTGCTA 661

QY 223 SerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLysIleCys 242
Db AGCCAAACAGAGAGCGAGCTGTCTGTACCAGGCAATCACGAAGAGTTGAAATATGT 721

QY 243 GluAsp 244
Db 722 GAAGAG 727

RESULT 6
AL536040

LOCUS
DEFINITION AL536040 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
CS0DF022Y005 5-PRIME, mRNA sequence.
ACCESSION AL536040
VERSION AL536040.2 GI:31260910
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT On Feb 13, 2001 this sequence version replaced gi:12799533.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7982.r For more information about this cluster, see <http://www.genoscope.cns.fr/>
cgi-bin/cluster.cgi?seq=CS0DF022AH03QPI&cluster=7982.r. Contact : Feng Liang Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DF022AH03QPI.

FEATURES
Location/Qualifiers
1. .1003
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF022Y005"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

BASE COUNT 308 a 221 c 273 g 196 t 5 others
ORIGIN

Alignment Scores:
Pred. No.: 2.16e-93 Length: 1003
Score: 1257.00 Matches: 240
Percent Similarity: 98.37% Conservative: 1


```
QY 161 LysArgAlaGlnGluSerProArgLysArgGlyArgProProLysAspGluLysAsp 180
    |||
    |||
    |||
Db 500 AAAAGAGCCCAAGAGCAAAAGTCCCGAAGCGGGGTGGGCCCAAGAGATGAGAAGGAT 559

QY 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
    |||
    |||
    |||
Db 560 CTCACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCTTT 619

QY 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
    |||
    |||
    |||
Db 620 AAATGGCAGCAACCGCAAGCAGCCTGTAAAGATGCAGATCCTCATTTCCATCATTTT 679

QY 221 LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeu--- 239
    |||
    |||
    |||
Db 680 CTGCTAAGCCCAACAGAGAGCCAGCTGTCTGCTACCAGGCCAATCACCGAAGAAGTTGA 739

QY 240 LysIleCysGlu 243
    |||
    |||
    |||
Db 740 AAATTATGTGAA 751
```

```
RESULT 8
BU520128
LOCUS
DEFINITION BU520128 993 bp mRNA linear EST 12-SEP-2002
IMAGE:6518047 5', mRNA sequence.
BU520128
VERSION BU520128.1 GI:22827654
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 993)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14099 row: c column: 08
High quality sequence stop: 675.
```

```
FEATURES
source
1..993
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6518047"
/tissue_type="undifferentiated limb"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 134"
/note="Vector: pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2:
NotI; Cloned unidirectionally. Primer: Oligo dt. Average
insert size 1.7 kb. Constructed by ResGen, Invitrogen
Corp. Note: this is a NIH MGC Library."
BASE COUNT 293 a 222 c 275 g 194 t 9 others
ORIGIN
```

```
Alignment Scores:
Pred. No.: 1.03e-91 Length: 993
Score: 1236.50 Matches: 233
Percent Similarity: 98.35% Conservative: 6
Best Local Similarity: 95.88% Mismatches: 3
Query Match: 93.75% Indels: 1
DB: 13 Gaps: 1
```

US-09-987-755-2 (1-249) x BU520128 (1-993)

```
QY 2 AlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProPro 21
    |||
    |||
    |||
Db 1 GCGGCTGTGAGTCTCGGCTCGCGACTTGGTGTGGGGAACCTGGGCCGTATCCTCCC 60

QY 22 TrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLysCys 41
    |||
    |||
    |||
Db 61 TGGCCAGGAAAGATTGTTAATCCACCCAGGACTTAAAGAAACCCACGTGGAAAGAAATGC 120

QY 42 PhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLys 61
    |||
    |||
    |||
Db 121 TTCTTTGTGAAGTTTTTTTGGAAACAGAAAGATCATGCTGGATCAAAGTGGAAAGCTAAAG 180

QY 62 ProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGln 81
    |||
    |||
    |||
Db 181 CCTTACCATGCTCACAGAGGAGATGATAAAGATTAAACAGGGTAAACCGTTCCAGCAA 240

QY 82 AlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSerSer 101
    |||
    |||
    |||
Db 241 GCTGTGGATGCTGTTGAAGAGTTCTCAGAGAGAGCCAAAGGGAAAGACCAGACATCATCC 300

QY 102 HisAsnSerSerAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsn 121
    |||
    |||
    |||
Db 301 CACACTTCTGCTGATGACAAGAAATCGCGTAATTCAGTGAGGAGAGAGTAGGCCCAAAC 360

QY 122 SerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGly 141
    |||
    |||
    |||
Db 361 TCAGGTGATGAGAAACGCAAGCTTAGCCTGTCTGAAGGAAAGGTGAAGAAGAACATGGGA 420

QY 142 GluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLys 161
    |||
    |||
    |||
Db 421 GAAGGAAAGAAAGAGGGTGACTTCAGGCTCTGCAGACAGAGGCTCCAAATGC---CTTAAA 477

QY 162 ArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAspLeu 181
    |||
    |||
    |||
Db 478 AGAGCCCAAGAGCAAAAGTCCCGAAGCGGGTGGGCCCAAGAGATGAGAAGGACCTC 537

QY 182 ThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLys 201
    |||
    |||
    |||
Db 538 ACCATCCCTGAGTCTAGCACTGTAAAGGGGATGATGGCTGGACCGATGGCTGCATTANA 597

QY 202 TrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPheLeu 221
    |||
    |||
    |||
Db 598 TGGCAGCCCAACAGCGACCGAGCCAGTCAAAGATGCAGATCCTCATTTTCATCTTCTG 657

QY 222 LeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLysIle 241
    |||
    |||
    |||
Db 658 TTGAGCCCAACAGAGAGCCAGCTGTCTGTACCAGGCAATCACAAAGAAAGTTGAAAATA 717

QY 242 CysGluAsp 244
    |||
    |||
    |||
Db 718 TGTGAAGAA 726
```

```
RESULT 9
BU520128
LOCUS
DEFINITION BU520128 890 bp mRNA linear EST 20-AUG-2002
AGENCOURT 8765993 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6330861
5', mRNA sequence.
ACCESSION BU520128
VERSION BU520128.1 GI:22332892
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 890)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
Ph.D.
```

cDNA Library Preparation: ResGen, Invitrogen Corp
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM13785 row: o column: 22
 High quality sequence stop: 648.

FEATURES

Location/Qualifiers

1. .890
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6330861"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC 130"
 /note="Organ: oocytes; Vector: pCMV-SPORT6.1.cdb;
 Site 1: EcoRV; Site 2: NotI; Cloned unidirectionally.
 Primer: Oligo dT. Average insert size 1.95 kb.
 Constructed by ResGen, Invitrogen Corp. Note: this is a
 NIH_MGC Library."

BASE COUNT 268 a 201 c 245 g 175 t 1 others
 ORIGIN

Alignment Scores:

Pred. No.: 1.97e-91 Length: 890
 Score: 1232.50 Matches: 232
 Percent Similarity: 98.35% Conservative: 6
 Best Local Similarity: 95.87% Mismatches: 3
 Query Match: 93.44% Indels: 1
 DB: 13 Gaps: 1

US-09-987-755-2 (1-249) x BQ918182 (1-890)

QY 3 AlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProProTrp 22
 |||||
 DB 3 GCTGTGAGTCTCGGCTCGCGACTGGTGTGGGAAACTGGCGGTATCCTCCCTGG 62
 QY 23 ProGlyLysIleValAsnProProLysAspLeuLysProArgGlyLysLysCysPhe 42
 |||||
 DB 63 CCAGAAAGATTGTTAATCCACCAGGACTTAAAGAAACCCAGTGGAAAGAAATGCTTC 122
 QY 43 PheValLysPheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysPro 62
 |||||
 DB 123 TTTGTGAAGTTTTTTGGAACAGAGATCATGCTGGATCAAAGTGAACAGCTAAAGCCT 182
 QY 63 TyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnAla 82
 |||||
 DB 183 TACCATGCTCACAGGAGGAGATGATTAAGATTAAACAAGGGTAAACGGTTCCAGCAAGCT 242
 QY 83 ValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSerSerHis 102
 |||||
 DB 243 GTGGATGCTGTTGAAGAGTTCTCAGGAGAGCCAAAGGGAAAGACAGACATATCCCAC 302
 QY 103 AsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsnSer 122
 |||||
 DB 303 ACTTCTGTGATGACAAAGATTCGGCGTAATTCAGTGAGGAGAGAGTAGGCCAAACTCA 362
 QY 123 GlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGlu 142
 |||||
 DB 363 GGTGATGAGAAACGCAAGCTTAGCTGTCTGAAGGGAAGGTGAAGAAGAACATGGGAGAA 422
 QY 143 GlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArg 162
 |||||
 DB 423 GGAAGAAGAGGGTGACTTCAGGCTCTGCAGACAGAGGCTCCAAATGC---CTTAAAGA 479
 QY 163 AlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAspLeuThr 182
 |||||
 DB 480 GCCAAGAGCAAAAGTCCCGGAAGGGGGTCCGGCCCCCAAGGATGAGAGGACCTCACC 539
 QY 183 IleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLysTrp 202
 |||||
 DB 540 ATCCCTGAGTCTAGCACTGTAAAGGGGATGATGGCTGGACCGATGCTGCTTAAATGG 599

QY 203 GlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPheLeuLeu 222
 |||||
 DB 600 CAGCCAAACAGCAGCCGAGTCAAAGATCAGATCCTCATTTTCATCATTTCTGTTG 659
 QY 223 SerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLysIleCys 242
 |||||
 DB 660 AGCCAAACAGAGAGCCAGCTGTCTGTACCANGCAATCACAAAGAGTTGAAATATGT 719
 QY 243 GluAsp 244
 |||||
 DB 720 GAAGAG 725

RESULT 10

BQ892009

LOCUS

DEFINITION

5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 899)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

Tissue Procurement: Susan L. Sullivan, Ph.D.

cDNA Library Preparation: ResGen, Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM13880 row: a column: 18

High quality sequence stop: 683.

Location/Qualifiers

1. .899

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="IMAGE:6391577"

/lab_host="DH10B (phage-resistant)"

/note="Organ: olfactory epithelium; Vector:

pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2: NotI; Cloned

unidirectionally. Primer: Oligo dT. Average insert size

2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this

is a NIH_MGC Library."

BASE COUNT 277 a 190 c 257 g 174 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 4.23e-91 Length: 899

Score: 1228.50 Matches: 231

Percent Similarity: 97.95% Conservative: 8

Best Local Similarity: 94.67% Mismatches: 4

Query Match: 93.14% Indels: 1

DB: 13 Gaps: 1

US-09-987-755-2 (1-249) x BQ892009 (1-899)

QY 1 MetAlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20

|||||

DB 5 ATGGCGGCTGTGAGTCTGGCGCTCGGCGACTGGTGTGGGGAACCTGGGCGGTATCCT 64

QY 21 ProTrpProGlyLysIleValAsnProProLysAspLeuLysProArgGlyLysLys 40

|||||

DB 65 CCCTGGCCAGGAAAGATTGTTAATCCACCAAGGACTTAAAGAAACCCACGCGGAAAGAA 124

```
Qy 41 CysPhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu 60
|||||
Db 125 TGCTTCTTTGTGAAGTTTTTTGGAAACAAGAATCATGCTGGATCAAAAGTGAACAGCTA 184

Qy 61 LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln 80
|||||
Db 185 AAGCCTTACCATGCTCAACAGGAGGAGATGATATAAGATTAAACAAAGGTAAACGGTTCCAG 244

Qy 81 GlnAlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSer 100
|||||
Db 245 CAAGCTGTGGATGCTGTGAAGAGTTCTCAGGAGAGCCAAAGGAAAGACCAGACATCA 304

Qy 101 SerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro 120
|||||
Db 305 TCCCACACTTCTGCTGATGACAAAGATCGCGTAATTCAGTGTGAGGAGAGTAGGCCA 364

Qy 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet 140
|||||
Db 365 AACTCAGTGTATGAGAAACGCAAGCTTAGCCTGTCTGAAGGGAGGTGAAGAAGAACATG 424

Qy 141 GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu 160
|||||
Db 425 GGAGAAAGGAAGAAGAGGCTGACTTCAGGCTCTGCAGACAGAGGCTCCAAATGC---CTT 481

Qy 161 LysArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp 180
|||||
Db 482 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTGGCGCCCCCAAGATGAGAAGGAG 541

Qy 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
|||||
Db 542 CTCACCATCCCTGAGTGTAGCACTGTAAAGGGGAGGATGGCTGGACCGATGGTGCAATT 601

Qy 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
|||||
Db 602 AAATGGCAGCCAAACAGCGACCGAGCCAGTCAAAAGATGCAGATCCTCAATTTTCATCATTTT 661

Qy 221 LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys 240
|||||
Db 662 CTGTTGAGCCGAACAGAGAAGCCAGCTGTCTGTACCAGCAATCACAAAGAAGTTGAAA 721

Qy 241 IleCysGluAsp 244
|||||
Db 722 ATATGTGAAGAG 733

RESULT 11
AK014456
LOCUS
DEFINITION
Mus musculus 13 days embryo whole body cDNA, RIKEN full-length
enriched library, clone:3930401K13 product:CYTOKINE-LIKE NUCLEAR
FACTOR N-PAC homolog [Homo sapiens], full insert sequence.
ACCESSION
AK014456
VERSION
AK014456.1 GI:12852319
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCE
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
JOURNAL
MEDLINE
PUBMED
11042159
REFERENCE
3
```

AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G.,
Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Sakamoto, N., Sasaki, H.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyo-Oka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohtsuki, S.
and Hayashizaki, Y.

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.

TITLE

JOURNAL

REFERENCE

AUTHORS

Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

TITLE

JOURNAL

REFERENCE

AUTHORS

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,
Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
Hiraoaka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,
Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K.,
Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE

JOURNAL

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

COMMENT

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
further details.

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGAGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA went

through one round of normalization to Rot = 10.0 and subtraction to Rot = 60.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATTCTCGAGTTAATTAATATCCCTCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: Sali; 3' end: BamHI. Host: DH10B.

FEATURES

source
Location/Qualifiers
1. .3239
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:3930401K13"
/db_xref="MGI:1907072"
/db_xref="taxon:10090"
/clone="3930401K13"
/tissue_type="whole body"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="13 days embryo"
8. .1648
/note="unnamed protein product; CYTOKINE-LIKE NUCLEAR FACTOR N-PAC homolog [Homo sapiens] (SPTR|Q9BXK2, evidence: FASTY, 97.8%ID, 100%length, match=1638) putative"
/codon_start=1
/protein_id="BAB29363.1"
/db_xref="GI:12852320"
/db_xref="MGI:1921272"
/translation="MAAVSLRLGLDVLWGKLGYPWPQKIVNPPKDLKPRGKKCFV

CDS

KFTGDKHAWIKVEQLPYHAHKEEMIKNGKRFFQAVDAVEEFLRAKGDQTSKH TSADDKWRNSSEERSRNSGDEKRLSLSEGVKKNMGEGKKRVTSGSADRGSKCLK RAQEQSPRRGRPPKDEKDLTIPESSTVKGMMAGPMFAFKQPTATPVPKADAPPHH FLLSQTERPVCYQAITKKLKICEEETGTSIQAADSTAVNGSITPTDKKIGFLGL MGSGIVNLLKMGHTVTWNRTAEKRGALGRTPAEVYSTCDITFACVSDPKAAKDLV LGPSGLVQIRPGKCYVDMSTVDADTVTELAQVIVSRGRFLEAPVSGNQQLSNDGML VILAAGDRLYEDCSSCFQAMGKTSFFLGEVGNAAKMLIVNMVQSGFMATIAEGLTL AQTGQSQOTLLDILNQQLASIFLDQKCNILQGNFKPDPFLYKLYIQKDLRLAIALGD AVNHPTPMAAANAENVYKRAKALDQSDNDMSAVYRAYIH"
BASE COUNT 832 a 785 c 837 g 785 t
ORIGIN

Alignment Scores:

Pred. No.: 2.65e-90 Length: 3239
Score: 1225.50 Matches: 231
Percent Similarity: 98.36% Conservative: 9
Best Local Similarity: 94.67% Mismatches: 3
Query Match: 92.91% Indels: 1
DB: 11 Gaps: 1

US-09-987-755-2 (1-249) x AK014456 (1-3239)

QY 1 MetAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
Db 8 ATGGCGGCTGTGAGTCTCGGCTCGCGGACTTGGTGTGGGGAACTGGCGCGGTATCCT 67
QY 21 ProTrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLys 40
Db 68 CCCTGCCAGGAAAGATTGTTAATCCACCCAGGACTTAAGAAACACGTCGGAAGAA 127
QY 41 CysPhePheValLysPheGlyThrGluAspHisAlaTyrIleLysValGluGlnLeu 60
Db 128 TGCCTCTTGTGAAGTTTGTGGAACAAAATCATGCTGGATCAAAAGTGAACAATA 187
QY 61 LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln 80
Db 188 AAGCCTTACCATTGCTCACAGGAGGAGATGATAAAGATTAAACAGGGTAAACGGTCCAG 247
QY 81 GlnAlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSer 100
Db 248 CAAGCTGTGGATGCTGTTGAAGAGTTCCTCAGGAGGCCAAAGGGAAGACCATCA 307
QY 101 SerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro 120

Db 308 TCCACACTTCTGCTGATGACAAGGATTGGCGTAATTCCAGTGAGGAGAGAAGTAGGCCA 367
QY 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet 140
Db 368 AACTCAGGTGATGAGAAACGCAAGCTTAGCCTGTCTGAAGGGAAGGTGAAGAAGACATG 427
QY 141 GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu 160
Db 428 GGAGAAGGAAAGAGGGGTGACTTCAGGCTCTGCAGACAGAGGCTCCAAATGC---CTT 484
QY 161 LysArgAlaGlnGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp 180
Db 485 AAAAGAGCCCAAGAGCAAAGTCCCGGAAGCGGGGTGCGGCCCCCAAAGGATGAGAAGGAC 544
QY 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
Db 545 CTCACCATCCCTGAGTCTAGCACTGTAAAGGGGATGATGGCTGGACCGATGGCTGCATTT 604
QY 201 LysTyrGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisHisPhe 220
Db 605 AAATGGCAGCCCAACAGCGACCGAGCCAGTCAAGATGCAGATCCTCATTTTTCATCATTTT 664
QY 221 LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys 240
Db 665 CTGTTGAGCCCAACAGAGAACCCAGCTGTCTGTACCAGGCAATCACAAAGAGTTGAAA 724
QY 241 IleCysGluAsp 244
Db 725 ATATGTGAAGAG 736

RESULT 12

BI551885
LOCUS 603192793F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5263991 5',
DEFINITION mRNA sequence.

ACCESSION BI551885

VERSION BI551885.1 GI:15439197

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 755)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki

Toshiyuki and Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM1665 row: j column: 24

High quality sequence stop: 746.

FEATURES

source

Location/Qualifiers
1. .755
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5263991"
/tissue_type="hippocampus".
/lab_host="DH10B"
/clone_lib="NIH_MGC_95"
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site 1: BamHI; Site 2: Sali-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the

Db 302 ATCCACACTTCTGCTGATGACAAGAATCGGCGTAATTCCAGTGAGGAGAGAAGTAGGCC 361

QY 120 oAsnSerGlyAspGluLysArgLysLeuSerLeuGluGlyLysValLysLysAsnMe. 140
|||||
Db 362 AAACCTCAGGTGATGAGAAACGCAAGCTTAGCCTGTCTGAAGGGAGGTGAAGAACAAT 421
QY 140 tGlyGluGlyLysLysArgValSerSerGlySerGluArgGlySerLysSerProLe 160
|||||
Db 422 GGGAGAAGGAAGAAGAGGGTGACTTCAGGCTCTGCAGACAGAGGCTCCAAATGC---CT 478

QY 160 uLysArgAlaGlnGlnSerProArgLysArgGlyArgProProLysAspGluLysAs 180
|||||
Db 479 TAAAGAGCCCAAGAGCAAGTCCCGGAAGCGGGTCGGCCCCCAAGAGGATGAGAAGGA 538

QY 180 pLeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPh 200
|||||
Db 539 CCTCACCATCCCTGAGTCTAGCACTGTAAAGGGGATGATGGCTGGACCGATGGCTGCATT 598

QY 200 eLysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisHisPh 220
|||||
Db 599 TAAATGGCAGCCCAACAGCGACCGAGCCAGTCAAAAGATGCAGATCCTCATTTTCATCATT 658

QY 220 eLeuLeuSer-GlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuL 240
|||||
Db 659 TCTGTTGAGCCCAACAGAGAGAGCCAGCTGTCTGTACCAGGCAATCACAAAGAAGTTGA 718

QY 240 yslleCysGlu 243
|||||
Db 719 AAATATGTGAA 729

RESULT 14

BQ934274 929 bp mRNA linear EST 21-AUG-2002
LOCUS AGENCOURT_8752717 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6335698
DEFINITION 5', mRNA sequence.

ACCESSION BQ934274

VERSION BQ934274.1 GI:22349657

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 929)

NIH-MGC <http://mgi.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman, Ph.D.

cDNA Library Preparation: ResGen, Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM13798 row: i column: 11

High quality sequence stop: 701.

FEATURES

source

1..929
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6335698"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_130"
/note="Organ: otocysts; Vector: pCMV-SPORT6.1.ccdB;
Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally.
Primer: Oligo dt. Average insert size 1.95 kb.
Constructed by ResGen, Invitrogen Corp. Note: this is a
NIH_MGC Library."

RESULT 15
AK040841
LOCUS AK040841 3250 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus adult male aorta and vein cDNA, RIKEN full-length enriched library, clone:A530029N17 product:CYTOKINE-LIKE NUCLEAR FACTOR N-PAC homolog [Homo sapiens], full insert sequence.

ACCESSION AK040841
VERSION AK040841.1 GI:26088102
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1

BASE COUNT 285 a 198 c 258 g 182 t 6 others

ORIGIN

Alignment Scores:
Pred. No.: 6 17e-90 Length: 929
Score: 1214.50 Matches: 228
Percent Similarity: 97.91% Conservative: 6
Best local Similarity: 95.40% Mismatches: 4
Query Match: 92.08% Indels: 1
DB: 13 Gaps: 1

US-09-987-755-2 (1-249) x BQ934274 (1-929)

QY 6 leuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProProTrpProGlyLys 25
|||||
Db 2 CTGCGGCTCGCGGACTTGGTGTGGGGAAACTGGGCCGTATCCTCCCTGGCCAGGAAAG 61

QY 26 ileValAsnProProLysAspLeuLysLysProArgGlyLysLysCysPheValLys 45
|||||
Db 62 ATTGTTAATCCACCCCAAGGACTTAAAGAAACACACGTGGAAGAAATGCTTCTTTGTGAAG 121

QY 46 PhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysProTyrHisAla 65
|||||
Db 122 TTTTGTGAACAGAGATCATGCTGGATCAAGTGAACAGTGAACAGTAAAGCCTTACCATGCT 181

QY 66 HisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGlnAlaValAspAla 85
|||||
Db 182 CACAAGGAGGAGATGATAAAGATTAAACAAGGTAACCGTTCCAGCAAGCTGTGGATGCT 241

QY 86 ValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSerSerHisAsnSerSer 105
|||||
Db 242 GTTGAAGAGTTCTCTCAGGAGAGCCAAAGGGAAGACCAAGACATCATCCACACTTCTGCT 301

QY 106 AspAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsnSerGlyAspGlu 125
|||||
Db 302 GATGACAAGATCGGCGTAATTCCAGTGAGGAGAGAAGTAGGCCAAACTCAGGTGATGAG 361

QY 126 LysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGlyLysLys 145
|||||
Db 362 AAACGCAAGCTTAGCCTGTCTGAAGGGAAGGTGAAGAAACATGGGAGAGGAAAGAG 421

QY 146 ArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArgAlaGlnGlu 165
|||||
Db 422 AGGTGACTTCAGGCTCTGCAGACAGAGGCTCCAAATGC---CTTAAAGAGGCCCAAGAG 478

QY 166 GlnSerProArgLysArgGlyArgProProLysAspGluLysAspLeuThrIleProGlu 185
|||||
Db 479 CAAAGTCCCCGGAAGCGGGTGGCCCCCAAGGATGAGAAGGACCTCACCATCCCTGAG 538

QY 186 SerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLysTyrGlnProThr 205
|||||
Db 539 TCTAGCACTGTAAAGGGGATGATGGTGGACCGATGGCTGATTTANATGGCAGCCAACA 598

QY 206 AlaSerGluProValLysAspAlaAspProHisPheHisHisPheLeuLeuSerGlnThr 225
|||||
Db 599 GCGACCGAGCCAGTCAAGATGCAGATCCTCATTTTTCATCATTTTCTGTTGAGCCAACA 658

QY 226 GluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLysIleCysGluAsp 244
|||||
Db 659 GAGAAGCCAGCTGTCTGTTACCAGGCAATCACAAAGAAAGTTGAANATATGTGAAGAG 715


```
QY 161 LysArgAlaGlnGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp 180
Db |||||
495 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGGATGAGAAGGAC 554
QY 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
Db |||||
555 CTCACCATCCCTGAGTCTAGCACTGTAAAGGGGATGATGGCTGGACCCGATGGCTGCATTT 614
QY 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
Db |||||
615 AAATGGCAGCCCAACAGCGACCGAGCCAGTCAAAGATGCAGATCCTCATTTTTCATCATTTT 674
QY 221 LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys 240
Db |||||
675 CTGTTGAGCCCAACAGAGAGAGCCAGCTGTCTGTTACCAGGCAATCACAAAGAAGTTGAAA 734
QY 241 IleCysGluAsp 244
Db |||||
735 ATATGTGAAGAG 746
```

Search completed: December 8, 2003, 13:25:41
Job time : 2159 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 8, 2003, 13:25:45 ; Search time 304 Seconds
(without alignments)
8166.870 Million cell updates/sec

Title: US-09-987-755-1_COPY_34_780
Perfect score: 747
Sequence: 1 atggcggctgtgagtcgtcg.....aagacctctcttcctagg 747

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2201672 seqs, 1661799599 residues

Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	747	100.0	3202	11	US-09-987-755-1
2	747	100.0	3256	15	US-10-103-313-190
3	745	99.7	3172	15	US-10-103-313-33
4	731	97.9	3727	13	US-10-067-482-1
5	722.4	96.7	3707	15	US-10-103-313-133
6	719.6	96.3	810	15	US-10-103-313-256
7	53.8	7.2	865	15	US-10-198-846-6497
8	53.8	7.2	872	15	US-10-198-846-6652
9	53.8	7.2	890	15	US-10-198-846-9404
10	53.8	7.2	897	15	US-10-198-846-6712
11	53.8	7.2	901	15	US-10-198-846-6624
12	53.8	7.2	942	15	US-10-198-846-6578
13	53.8	7.2	1121	15	US-10-198-846-8409
C 14	53.2	7.1	822	15	US-10-198-846-9427
C 15	53.2	7.1	863	15	US-10-198-846-6557

16	52.8	7.1	842	15	US-10-198-846-4084	Sequence 4084, Ap
17	52.8	7.1	1127	15	US-10-198-846-8540	Sequence 8540, Ap
C 18	51.6	6.9	1038	15	US-10-198-846-9410	Sequence 9410, Ap
19	47	6.3	2920	11	US-09-866-050A-541	Sequence 541, App
20	47	6.3	2920	15	US-10-152-661-541	Sequence 541, App
21	44	5.9	1898	9	US-09-768-826-22	Sequence 22, Appl
22	44	5.9	3805	13	US-10-247-671-8	Sequence 8, Appl
23	41	5.5	288	10	US-09-960-352-1025	Sequence 1025, Ap
24	40.4	5.4	594	13	US-10-140-472-10	Sequence 10, Appl
25	40.4	5.4	594	13	US-10-141-761-10	Sequence 10, Appl
26	40.4	5.4	594	13	US-10-142-885-10	Sequence 10, Appl
27	40.4	5.4	594	13	US-10-158-790-10	Sequence 10, Appl
28	40.4	5.4	594	13	US-10-137-871-10	Sequence 10, Appl
29	40.4	5.4	594	13	US-10-140-805-10	Sequence 10, Appl
30	40.4	5.4	594	13	US-10-140-864-10	Sequence 10, Appl
31	40.4	5.4	594	13	US-10-140-923-10	Sequence 10, Appl
32	40.4	5.4	594	13	US-10-141-756-10	Sequence 10, Appl
33	40.4	5.4	594	13	US-10-141-759-10	Sequence 10, Appl
34	40.4	5.4	594	15	US-10-123-155-10	Sequence 10, Appl
35	40.4	5.4	594	16	US-10-146-731-10	Sequence 10, Appl
C 36	37.2	5.0	491	10	US-09-796-692-8648	Sequence 8648, Ap
C 37	37.2	5.0	491	15	US-10-040-862-8648	Sequence 8648, Ap
38	37.2	5.0	1919	9	US-09-925-301-145	Sequence 145, App
39	37.2	5.0	4866	13	US-10-007-926A-83	Sequence 83, Appl
40	37.2	5.0	4866	15	US-10-171-581-140	Sequence 140, App
C 41	36.4	4.9	552	15	US-10-198-846-6607	Sequence 6607, Ap
42	36	4.8	1297	9	US-09-768-826-24	Sequence 24, Appl
43	36	4.8	2031	11	US-09-823-187-23	Sequence 23, Appl
44	36	4.8	2272	11	US-09-946-374-307	Sequence 307, App
45	36	4.8	2272	13	US-10-015-387A-307	Sequence 307, App

ALIGNMENTS

RESULT 1
US-09-987-755-1
; Sequence 1, Application US/09987755
; Publication No. US20030022312A1
; GENERAL INFORMATION:
; APPLICANT: Kunsch et al
; TITLE OF INVENTION: Human Hepatoma-Derived Growth Factor-2
; FILE REFERENCE: PF198D1C1
; CURRENT APPLICATION NUMBER: US/09/987,755
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 09/263,625
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: 08/464,600
; PRIOR FILING DATE: 1995-06-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 3202
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (34)..(780)
; OTHER INFORMATION:
US-09-987-755-1

Query Match 100.0%; Score 747; DB 11; Length 3202;
Best Local Similarity 100.0%; Pred. No. 3.9e-227;
Matches 747; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGGCGGCTGTGAGTCTCGGCTCGGCGGAGTGTGGTGGGAACTCGGCCGATATCCT	60
Db	34	ATGGCGGCTGTGAGTCTCGGCTCGGCGGAGTGTGGTGGGAACTCGGCCGATATCCT	93
QY	61	CCTGGCCAGGAAGATTGTTATCCACCAAGAACTTGAAGAACTCGCGGAAAGAAA	120
Db	94	CCTGGCCAGGAAGATTGTTATCCACCAAGAACTTGAAGAACTCGCGGAAAGAAA	153

QY 121 TGCTTCTTTGTGAAATTTTGGAAACAGAGATCATGCTCGGATCAAAAGTGAACAGCTG 180
Db 154 TGCTTCTTTGTGAAATTTTGGAAACAGAGATCATGCTCGGATCAAAAGTGAACAGCTG 213
QY 181 AAGCCATATCATGCTCATAAAGAGGAATGATAAAATTAACAAGGTAAACGATTCAG 240
Db 214 AAGCCATATCATGCTCATAAAGAGGAATGATAAAATTAACAAGGTAAACGATTCAG 273
QY 241 CAAGCGGTAGATGCTGTCGAAGAGTTCTCAGGAGAGCCAAAGGAAAGACAGACGTC 300
Db 274 CAAGCGGTAGATGCTGTCGAAGAGTTCTCAGGAGAGCCAAAGGAAAGACAGACGTC 333
QY 301 TCCCACAATTTCTTGATGACAAGAAATCGACGTAATTCAGTGAGGAGAAAGTGGCCA 360
Db 334 TCCCACAATTTCTTGATGACAAGAAATCGACGTAATTCAGTGAGGAGAAAGTGGCCA 393
QY 361 AACTCAGGTGATGAGAAAGCGCAAACTTAGCCTGTCTGAAGGAAAGTGAAGAAACATG 420
Db 394 AACTCAGGTGATGAGAAAGCGCAAACTTAGCCTGTCTGAAGGAAAGTGAAGAAACATG 453
QY 421 GGAGAAAGGAAAGAGAGGAGTCTTTCAGGCTCTTTCAGGAGAGGCTCCAAATCCCCTCTG 480
Db 454 GGAGAAAGGAAAGAGAGGAGTCTTTCAGGCTCTTTCAGGAGAGGCTCCAAATCCCCTCTG 513
QY 481 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGAGTGAAGAGAT 540
Db 514 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGAGTGAAGAGAT 573
QY 541 CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGCGCGGACCGATGGCCGCTT 600
Db 574 CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGCGCGGACCGATGGCCGCTT 633
QY 601 AAATGGCAGCAACCGCAAGCGAGCCTGTAAAGATGCAATCCTCATTTCCATTTTC 660
Db 634 AAATGGCAGCAACCGCAAGCGAGCCTGTAAAGATGCAATCCTCATTTCCATTTTC 693
QY 661 CTGCTAAGCCAAACAGAGAGCCAGCTGTCTTACCAGCAATCAGAAAGTGGAAA 720
Db 694 CTGCTAAGCCAAACAGAGAGCCAGCTGTCTTACCAGCAATCAGAAAGTGGAAA 753
QY 721 ATATGTGAAGACCTCTTCTTCTTAGG 747
Db 754 ATATGTGAAGACCTCTTCTTCTTAGG 780

RESULT 2
US-10-103-313-190
; Sequence 190, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJZ07C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 190
; LENGTH: 3256
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3220)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-103-313-190
Query Match 100.0%; Score 747; DB 15; Length 3256;
Best Local Similarity 100.0%; Pred. No. 4e-227;
Matches 747; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCGGCTGTGAGTCTCGGCTCGGAGTGGTGTGGGGAAACTCGGCCGATATCCT 60

Db 34 ATGGCGGCTGTGAGTCTCGGCTCGGAGTGGTGTGGGGAAACTCGGCCGATATCCT 93
QY 61 CCTTGGCCAGGAAAGATTTGTAATCCACCAAGAGCTTGAAGAAACCTCGCGGAAAGAAA 120
Db 94 CCTTGGCCAGGAAAGATTTGTAATCCACCAAGAGCTTGAAGAAACCTCGCGGAAAGAAA 153
QY 121 TGCTTCTTTGTGAAATTTTGGAAACAGAGATCATGCTCGGATCAAAAGTGAACAGCTG 180
Db 154 TGCTTCTTTGTGAAATTTTGGAAACAGAGATCATGCTCGGATCAAAAGTGAACAGCTG 213
QY 181 AAGCCATATCATGCTCATAAAGAGGAATGATAAAATTAACAAGGTAAACGATTCAG 240
Db 214 AAGCCATATCATGCTCATAAAGAGGAATGATAAAATTAACAAGGTAAACGATTCAG 273
QY 241 CAAGCGGTAGATGCTGTCGAAGAGTTCTCAGGAGAGCCAAAGGAAAGACAGACGTC 300
Db 274 CAAGCGGTAGATGCTGTCGAAGAGTTCTCAGGAGAGCCAAAGGAAAGACAGACGTC 333
QY 301 TCCCACAATTTCTTGATGACAAGAAATCGACGTAATTCAGTGAGGAGAAAGTGGCCA 360
Db 334 TCCCACAATTTCTTGATGACAAGAAATCGACGTAATTCAGTGAGGAGAAAGTGGCCA 393
QY 361 AACTCAGGTGATGAGAAAGCGCAAACTTAGCCTGTCTGAAGGAAAGTGAAGAAACATG 420
Db 394 AACTCAGGTGATGAGAAAGCGCAAACTTAGCCTGTCTGAAGGAAAGTGAAGAAACATG 453
QY 421 GGAGAAAGGAAAGAGAGGAGTCTTTCAGGCTCTTTCAGAGAGAGGCTCCAAATCCCCTCTG 480
Db 454 GGAGAAAGGAAAGAGAGGAGTCTTTCAGGCTCTTTCAGAGAGAGGCTCCAAATCCCCTCTG 513
QY 481 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGAGTGAAGAGAT 540
Db 514 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGAGTGAAGAGAT 573
QY 541 CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGCGCGGACCGATGGCCGCTT 600
Db 574 CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGCGCGGACCGATGGCCGCTT 633
QY 601 AAATGGCAGCAACCGCAAGCGAGCCTGTAAAGATGCAATCCTCATTTCCATTTTC 660
Db 634 AAATGGCAGCAACCGCAAGCGAGCCTGTAAAGATGCAATCCTCATTTCCATTTTC 693
QY 661 CTGCTAAGCCAAACAGAGAGCCAGCTGTCTTACCAGCAATCAGAAAGTGGAAA 720
Db 694 CTGCTAAGCCAAACAGAGAGCCAGCTGTCTTACCAGCAATCAGAAAGTGGAAA 753
QY 721 ATATGTGAAGACCTCTTCTTCTTAGG 747
Db 754 ATATGTGAAGACCTCTTCTTCTTAGG 780

RESULT 3
US-10-103-313-33
; Sequence 33, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJZ07C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 3172
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-103-313-33
Query Match 99.7%; Score 745; DB 15; Length 3172;
Best Local Similarity 100.0%; Pred. No. 1.7e-226;


```
Matches 745; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GCGGCTGTGAGTCTGCGGCTCGCGGACCTGGTGTGGGAAACTCGCGGATATCCTCC 62
Db |||||
QY 8 GCGGCTGTGAGTCTGCGGCTCGCGGACCTGGTGTGGGAAACTCGCGGATATCCTCC 67
Db |||||
QY 63 TTGGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAATG 122
Db |||||
QY 68 TTGGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAATG 127
Db |||||
QY 123 CTTCTTTGTGAAATTTTGGAAACAGAAAGATCATGCTGGATCAAAAGTGGAACAGCTGAA 182
Db |||||
QY 128 CTTCTTTGTGAAATTTTGGAAACAGAAAGATCATGCTGGATCAAAAGTGGAACAGCTGAA 187
Db |||||
QY 183 GCCATATCATGCTCATAAAGAGGAAATGATAAAATTAACAAGGTTAAACGATTCCAGCA 242
Db |||||
QY 188 GCCATATCATGCTCATAAAGAGGAAATGATAAAATTAACAAGGTTAAACGATTCCAGCA 247
Db |||||
QY 243 AGCGGTAGATGCTGTCGAAGAGTTCTCAGGAGAGCCAAAGGAAAGACCCAGACGTATC 302
Db |||||
QY 248 AGCGGTAGATGCTGTCGAAGAGTTCTCAGGAGAGCCAAAGGAAAGACCCAGACGTATC 307
Db |||||
QY 303 CCACAATTTCTTGATGACAAGAAATCGACGTAATCCAGTGAGGAGAGAGTAGGCCAAA 362
Db |||||
QY 308 CCACAATTTCTTGATGACAAGAAATCGACGTAATCCAGTGAGGAGAGAGTAGGCCAAA 367
Db |||||
QY 363 CTCAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGAAGGGAAGTGAAGAAACATGGG 422
Db |||||
QY 368 CTCAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGAAGGGAAGTGAAGAAACATGGG 427
Db |||||
QY 423 AGAAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAA 482
Db |||||
QY 428 AGAAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAA 487
Db |||||
QY 483 AAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTGGGCCCCCAAGGATGAGAGGATCT 542
Db |||||
QY 488 AAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTGGGCCCCCAAGGATGAGAGGATCT 547
Db |||||
QY 543 CACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGGCCGGACCGATGGCCGGTTTAA 602
Db |||||
QY 548 CACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGGCCGGACCGATGGCCGGTTTAA 607
Db |||||
QY 603 ATGCAGCCCAACCGCAAGCGAGCTGTAAAGATGCAGATCCTCATTTCCATCTTCT 662
Db |||||
QY 608 ATGCAGCCCAACCGCAAGCGAGCTGTAAAGATGCAGATCCTCATTTCCATCTTCT 667
Db |||||
QY 663 GCTAAGCCCAACAGAGAAAGCCAGCTGTGTTACAGGCAATACGAAGAAAGTTGAAAT 722
Db |||||
QY 668 GCTAAGCCCAACAGAGAAAGCCAGCTGTGTTACAGGCAATACGAAGAAAGTTGAAAT 727
Db |||||
QY 723 ATGTGAAGACCTCCTTCTCCTAGG 747
Db |||||
QY 728 ATGTGAAGACCTCCTTCTCCTAGG 752
Db |||||
```

RESULT 4

```
US-10-067-482-1
; Sequence 1, Application US/10067482
; Publication No. US20030148407A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc.
; TITLE OF INVENTION: Angiogenesis Dehydrogenase Gene
; FILE REFERENCE: 1U 102 R1
; CURRENT APPLICATION NUMBER: US/10/067,482
; CURRENT FILING DATE: 2002-02-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 3727
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: CDS
```

```
; LOCATION: (28)..(1689)
; OTHER INFORMATION:
US-10-067-482-1
```

Query Match 97.9%; Score 731; DB 13; Length 3727;
Best Local Similarity 100.0%; Pred. No. 5.4e-222;
Matches 731; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGGCGGCTGTGAGTCTGCGGCTCGCGGACTTGGTGTGGGGAAACTCGCGGATATCCT 60
Db |||||
QY 28 ATGGCGGCTGTGAGTCTGCGGCTCGCGGACTTGGTGTGGGGAAACTCGCGGATATCCT 87
Db |||||
QY 61 CCTTGGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAA 120
Db |||||
QY 88 CCTTGGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAA 147
Db |||||
QY 121 TGCTTCTTTGTGAAATTTTGGAAACAGAAAGATCATGCTGGATCAAAAGTGGAACAGCTG 180
Db |||||
QY 148 TGCTTCTTTGTGAAATTTTGGAAACAGAAAGATCATGCTGGATCAAAAGTGGAACAGCTG 207
Db |||||
QY 181 AAGCCATATCATGCTCATAAAGAGGAAATGATAAAATTAACAAGGTTAAACGATTCCAG 240
Db |||||
QY 208 AAGCCATATCATGCTCATAAAGAGGAAATGATAAAATTAACAAGGTTAAACGATTCCAG 267
Db |||||
QY 241 CAAGCGGTAGATGCTGTCGAAGAGTTCTCAGGAGAGCCAAAGGAAAGACCCAGACGTCA 300
Db |||||
QY 268 CAAGCGGTAGATGCTGTCGAAGAGTTCTCAGGAGAGCCAAAGGAAAGACCCAGACGTCA 327
Db |||||
QY 301 TCCACAAATTTCTGATGACAAGAAATCGACGTAATTCAGTGAGGAGAGAGTAGGCCCA 360
Db |||||
QY 328 TCCACAAATTTCTGATGACAAGAAATCGACGTAATTCAGTGAGGAGAGAGTAGGCCCA 387
Db |||||
QY 361 AACTCAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAACATG 420
Db |||||
QY 388 AACTCAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAACATG 447
Db |||||
QY 421 GGAGAAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG 480
Db |||||
QY 448 GGAGAAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG 507
Db |||||
QY 481 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTGGGCCCCCAAGGATGAGAAGGAT 540
Db |||||
QY 508 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTGGGCCCCCAAGGATGAGAAGGAT 567
Db |||||
QY 541 CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGGACCGATGGCCGGTTT 600
Db |||||
QY 568 CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGGACCGATGGCCGGTTT 627
Db |||||
QY 601 AAATGGCAGCAACCGCAAGCGGACCTGTTAAAGATGCAGATCCTCATTTCCATCATTT 660
Db |||||
QY 628 AAATGGCAGCAACCGCAAGCGGACCTGTTAAAGATGCAGATCCTCATTTCCATCATTT 687
Db |||||
QY 661 CTGCTAAGCCAAACAGAGAACGACGCTGTGTTACAGGCAATACGAAGAAAGTTGAAA 720
Db |||||
QY 688 CTGCTAAGCCAAACAGAGAACGACGCTGTGTTACAGGCAATACGAAGAAAGTTGAAA 747
Db |||||
QY 721 ATATGTGAAGA 731
Db |||||
QY 748 ATATGTGAAGA 758
Db |||||
```

RESULT 5

```
US-10-103-313-133
; Sequence 133, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ207C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 3707
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-103-313-133

Query Match      96.7%; Score 722.4; DB 15; Length 3707;
Best Local Similarity 99.9%; Pred. No. 3e-219;
Matches 723; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 8 CTGTGAGTCTGCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCCGATATCCTCTTGGC 67
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 18 CGGTGAGTCTGCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCCGATATCCTCTTGGC 77
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 68 CAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAATGCTTCT 127
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 78 CAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAATGCTTCT 137
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 128 TTGTGAAATTTTGTGAAACAGAGATCATGCTGGATCAAAAGTGAACACAGCTGAAGCCAT 187
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 138 TTGTGAAATTTTGTGAAACAGAGATCATGCTGGATCAAAAGTGAACACAGCTGAAGCCAT 197
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 188 ATCATGCTCATAAAGAGGAAATGATAAAATTAACAAGGTAACGATTCCAGCAAGCGG 247
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 198 ATCATGCTCATAAAGAGGAAATGATAAAATTAACAAGGTAACGATTCCAGCAAGCGG 257
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 248 TAGATGCTCTGAAGAGTTCTTCAGGAGAGCCAAAGGGAAGACCAGAGCTCATCCCACA 307
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 258 TAGATGCTCTGAAGAGTTCTTCAGGAGAGCCAAAGGGAAGACCAGAGCTCATCCCACA 317
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 308 ATTCTTCTGATGACAAGATCGACGTAATTCAGTGAGGAGAGAGTAGGCCAAACTCAG 367
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 318 ATTCTTCTGATGACAAGATCGACGTAATTCAGTGAGGAGAGAGTAGGCCAAACTCAG 377
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 368 GTGATGAGAGCGCAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAAGACATGGGAGAG 427
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 378 GTGATGAGAGCGCAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAAGACATGGGAGAG 437
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 428 GAAAGAAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTGAAAAGAG 487
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 438 GAAAGAAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTGAAAAGAG 497
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 488 CCCAAGAGCAAAAGTCCCAGGAGCGGGTCCGCCCCCAAGAGGATGAGAAGGATCTCACCA 547
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 498 CCCAAGAGCAAAAGTCCCAGGAGCGGGTCCGCCCCCAAGAGGATGAGAAGGATCTCACCA 557
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 548 TCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGGACCGATGGCCGGCTTTAAATGGC 607
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 558 TCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGGACCGATGGCCGGCTTTAAATGGC 617
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 608 AGCCAAACCGAAGCGAGCTGTTAAAGATGCAGATCCTCATTTCCATCATTTCTGCTAA 667
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 618 AGCCAAACCGAAGCGAGCTGTTAAAGATGCAGATCCTCATTTCCATCATTTCTGCTAA 677
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 668 GCCAAACAGAGAGCCAGCTGTCTGTTACAGGCAATCAAGAAAGAGTTGAAATATGTG 727
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 678 GCCAAACAGAGAGCCAGCTGTCTGTTACAGGCAATCAAGAAAGAGTTGAAATATGTG 737
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 728 AAGA 731
Db | | |
Qy 738 AAGA 741
Db | | |
```

RESULT 6

```
US-10-103-313-256
; Sequence 256, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ207C1
; CURRENT APPLICATION NUMBER: US/10/103,313

; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 256
; LENGTH: 810
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (12)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (43)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-103-313-256

Query Match      96.3%; Score 719.6; DB 15; Length 810;
Best Local Similarity 99.6%; Pred. No. 1e-218;
Matches 719; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 10 GTGAGTCTGCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCCGATATCCTCTTGGCCA 69
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 70 GGAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAATGCTTCTTT 129
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 62 GGAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAATGCTTCTTT 121
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 130 GTGAATTTTGGAAACAGAGATCATGCTGGATCAAAAGTGAACACAGCTGAAGCCATAT 189
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 122 GTGAATTTTGGAAACAGAGATCATGCTGGATCAAAAGTGAACACAGCTGAAGCCATAT 181
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 190 CATGCTCATAAAGAGGAAATGATAAAATTAACAAGGTAACGATTCCAGCAAGCGTA 249
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 182 CATGCTCATAAAGAGGAAATGATAAAATTAACAAGGTAACGATTCCAGCAAGCGTA 241
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 250 GATGCTGCGAAGAGTTCTTCAGGAGAGCCAAAGGGAAGACCAGAGCTCATCCCACAT 309
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 242 GATGCTGCGAAGAGTTCTTCAGGAGAGCCAAAGGGAAGACCAGAGCTCATCCCACAT 301
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 310 TCTTCTGATGACAAGATCGACGTAATTCAGTGAGGAGAGAGTAGGCCAAACTCAGGT 369
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 302 TCTTCTGATGACAAGATCGACGTAATTCAGTGAGGAGAGAGTAGGCCAAACTCAGGT 361
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 370 GATGAGAAGCGCAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAAGACATGGGAGAGGA 429
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 362 GATGAGAAGCGCAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAAGACATGGGAGAGGA 421
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 430 AAGAAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTGAAAAGAGCC 489
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 422 AAGAAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTGAAAAGAGCC 481
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 490 CAAGAGCAAAAGTCCCAGGAAGCGGGTCCGCCCCCAAGAGGATGAGAAGGATCTCACCATC 549
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 482 CAAGAGCAAAAGTCCCAGGAAGCGGGTCCGCCCCCAAGAGGATGAGAAGGATCTCACCATC 541
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 550 CCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGGACCGATGGCCGGCTTTAAATGGCAG 609
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 542 CCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGGACCGATGGCCGGCTTTAAATGGCAG 601
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 610 CCAACCGCAAGCGAGCTGTTAAAGATGCAGATCCTCATTTCCATCATTTCTGCTAAGC 669
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 602 CCAACCGCAAGCGAGCTGTTAAAGATGCAGATCCTCATTTCCATCATTTCTGCTAAGC 661
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 670 CAAACAGAGAGAGCCAGCTGTCTGTTACCAGGCAATCAAGAAAGAGTTGAAATATGTGAA 729
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 662 CAAACAGAGAGAGCCAGCTGTCTGTTACCAGGCAATCAAGAAAGAGTTGAAATATGTGAA 721
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 730 GA 731
Db | |
Qy 722 GA 723
Db | |
```

```
RESULT 7
US-10-198-846-6497
; Sequence 6497, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6497
; LENGTH: 865
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 262, 406, 416, 417, 437, 449, 451, 455, 465, 474, 494, 498,
; LOCATION: 515, 517, 550, 584, 587, 607, 619, 620, 626, 652, 662, 663,
; LOCATION: 666, 691, 697, 711, 722, 728, 764, 776, 777, 802, 804, 810,
; LOCATION: 812, 819, 822, 824, 825, 826, 829, 831, 836, 848, 849
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 850, 856, 860, 862, 864
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-6497

Query Match          7.2%; Score 53.8; DB 15; Length 865;
Best Local Similarity 89.2%; Pred. No. 2.7e-06;
Matches 58; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAACCTCGGCCGATATCCT 60
Db 67 ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAACCTCGGCCGCTCTAGA 126

QY 61 CCTTG 65
Db 127 ACTAG 131

RESULT 8
US-10-198-846-6652
; Sequence 6652, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6652
; LENGTH: 872
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 455, 590, 626, 634, 702, 703, 766, 772, 775, 787, 803, 810,
; LOCATION: 814, 821, 828, 835, 850, 852, 855, 872
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-6652

Query Match          7.2%; Score 53.8; DB 15; Length 872;
Best Local Similarity 89.2%; Pred. No. 2.7e-06;
Matches 58; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAACCTCGGCCGATATCCT 60
Db 58 ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAACCTCGGCCGCTCTAGA 117

QY 61 CCTTG 65
Db 118 ACTAG 122

RESULT 9
US-10-198-846-9404
; Sequence 9404, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9404
; LENGTH: 890
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 515, 565, 609, 617, 623, 633, 637, 648, 658, 659, 703, 709,
; LOCATION: 719, 723, 727, 742, 749, 753, 755, 757, 763, 767, 768, 770,
; LOCATION: 774, 778, 779, 781, 782, 783, 786, 790, 792, 799, 802, 806,
; LOCATION: 814, 815, 819, 825, 829, 836, 838, 849, 850, 858, 865
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 868, 873, 874, 877, 878, 879, 881, 885, 887, 889, 890
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-9404

Query Match          7.2%; Score 53.8; DB 15; Length 890;
Best Local Similarity 89.2%; Pred. No. 2.7e-06;
Matches 58; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAACCTCGGCCGATATCCT 60
Db 65 ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAACCTCGGCCGCTCTAGA 124

QY 61 CCTTG 65
Db 125 ACTAG 129

RESULT 10
US-10-198-846-6712
; Sequence 6712, Application US/10198846
; Publication No. US20030099974A1
```

GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Xu, Yongyao
APPLICANT: Wang, Youzhen
APPLICANT: Steinmann, Kathleen
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
TITLE OF INVENTION: THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-049
CURRENT APPLICATION NUMBER: US/10/198,846
CURRENT FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: 60/306,220
PRIOR FILING DATE: 2001-07-18
NUMBER OF SEQ ID NOS: 14084
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6712
LENGTH: 897
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 141, 178, 247, 310, 405, 423, 438, 440, 445, 446, 452, 453,
LOCATION: 487, 490, 492, 503, 515, 517, 522, 531, 561, 563, 564, 587,
LOCATION: 599, 604, 623, 643, 650, 676, 720, 721, 727, 729, 736, 744,
LOCATION: 747, 749, 752, 755, 758, 764, 766, 770, 781, 786, 792
OTHER INFORMATION: n = A,T,C or G
FEATURE:
NAME/KEY: misc_feature
LOCATION: 799, 800, 821, 826, 831, 836, 839, 843, 850, 854, 866, 870,
LOCATION: 874, 879, 884, 888
OTHER INFORMATION: n = A,T,C or G
US-10-198-846-6712
Query Match 7.2%; Score 53.8; DB 15; Length 897;
Best Local Similarity 89.2%; Pred. No. 2.8e-06;
Matches 58; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Qy 1 ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCCGATATCCT 60
Db 64 ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCCGCTCTAGA 123
Qy 61 CCTTG 65
Db 124 ACTAG 128
RESULT 11
US-10-198-846-6624
Sequence 6624, Application US/10198846
Publication No. US2003009974A1
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Xu, Yongyao
APPLICANT: Wang, Youzhen
APPLICANT: Steinmann, Kathleen
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
TITLE OF INVENTION: THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-049
CURRENT APPLICATION NUMBER: US/10/198,846
CURRENT FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: 60/306,220
PRIOR FILING DATE: 2001-07-18
NUMBER OF SEQ ID NOS: 14084
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6624
LENGTH: 901
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3, 4, 11, 13, 14, 33, 298, 299, 300, 301, 302, 303, 304,
LOCATION: 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316,

LOCATION: 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328,
LOCATION: 329, 330, 331, 332, 334, 335, 336, 337, 338, 339, 340
OTHER INFORMATION: n = A,T,C or G
FEATURE:
NAME/KEY: misc_feature
LOCATION: 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352,
LOCATION: 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364,
LOCATION: 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376,
LOCATION: 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387
OTHER INFORMATION: n = A,T,C or G
FEATURE:
NAME/KEY: misc_feature
LOCATION: 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399,
LOCATION: 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411,
LOCATION: 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423,
LOCATION: 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434
OTHER INFORMATION: n = A,T,C or G
FEATURE:
NAME/KEY: misc_feature
LOCATION: 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446,
LOCATION: 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458,
LOCATION: 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470,
LOCATION: 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481
OTHER INFORMATION: n = A,T,C or G
FEATURE:
NAME/KEY: misc_feature
LOCATION: 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493,
LOCATION: 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505,
LOCATION: 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517,
LOCATION: 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528
OTHER INFORMATION: n = A,T,C or G
FEATURE:
NAME/KEY: misc_feature
LOCATION: 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540,
LOCATION: 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552,
LOCATION: 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564,
LOCATION: 565-778, 780-851, 855-900
OTHER INFORMATION: n = A,T,C or G
US-10-198-846-6624
Query Match 7.2%; Score 53.8; DB 15; Length 901;
Best Local Similarity 89.2%; Pred. No. 2.8e-06;
Matches 58; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Qy 1 ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCCGATATCCT 60
Db 90 ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCCGCTCTAGA 149
Qy 61 CCTTG 65
Db 150 ACTAG 154
RESULT 12
US-10-198-846-6578
Sequence 6578, Application US/10198846
Publication No. US2003009974A1
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Xu, Yongyao
APPLICANT: Wang, Youzhen
APPLICANT: Steinmann, Kathleen
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
TITLE OF INVENTION: THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-049
CURRENT APPLICATION NUMBER: US/10/198,846
CURRENT FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: 60/306,220
PRIOR FILING DATE: 2001-07-18
NUMBER OF SEQ ID NOS: 14084
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6578

```
;
; LENGTH: 942
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 325, 393, 415, 458, 478, 506, 511, 518, 520, 525, 529, 531,
; LOCATION: 534, 539, 552, 567, 575, 593, 594, 596, 598, 600, 601, 602,
; LOCATION: 603, 608, 613, 618, 625, 631, 633, 635, 641, 643, 644, 645,
; LOCATION: 652, 653, 658, 660, 666, 669, 676, 693, 694, 699, 700
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 703, 705, 707, 714, 715, 719, 721, 722, 725, 726, 736, 737,
; LOCATION: 741, 753, 760, 761, 762, 772, 773, 775, 783, 785, 804, 805,
; LOCATION: 807, 809, 816, 817, 822, 826, 835, 838, 847, 848, 849, 850,
; LOCATION: 853, 854, 855, 856, 857, 863, 868, 869, 872, 875, 876
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 883, 885, 888, 889, 897, 899, 901, 904, 905, 908, 912, 915,
; LOCATION: 921, 923, 930, 937, 940
; OTHER INFORMATION: n = A,T,C or G
; US-10-198-846-6578

Query Match          7.2%; Score 53.8; DB 15; Length 942;
Best Local Similarity 89.2%; Pred. No. 2.8e-06;
Matches 58; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAACCTCGGCCGATATCCT 60
    |||||
Db 56 ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAACCTCGGCCGATATCCT 60
    |||||

QY 61 CCTTG 65
    |||
Db 116 ACTAG 120

RESULT 13
US-10-198-846-8409
; Sequence 8409, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8409
; LENGTH: 1121
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 3, 4, 6, 12, 13, 17, 22, 23, 27, 29, 32, 38, 39, 47, 48,
; LOCATION: 50, 51, 54, 62, 63, 66, 67, 70, 71, 72, 73, 75, 77, 79,
; LOCATION: 80, 81, 82, 84, 86, 87, 88, 90, 92, 93, 99, 100, 101, 110,
; LOCATION: 111, 297, 428, 481, 569, 685, 689, 736, 751, 756, 761, 764
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 774, 775, 781, 782, 792, 813, 848, 851, 857, 863, 868, 875,
; LOCATION: 882, 886, 911, 926, 927, 938, 940, 941, 942, 951, 952, 964,
; LOCATION: 966, 967, 971, 979, 985, 987, 997, 1004, 1032, 1034, 1050,
; LOCATION: 1062, 1064, 1092, 1097, 1099, 1107, 1114, 1115, 1116
```

```
;
; OTHER INFORMATION: n = A,T,C or G
; US-10-198-846-8409

Query Match          7.2%; Score 53.8; DB 15; Length 1121;
Best Local Similarity 89.2%; Pred. No. 3.1e-06;
Matches 58; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAACCTCGGCCGATATCCT 60
    |||||
Db 178 ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAACCTCGGCCGCTCTAGA 237
    |||||

QY 61 CCTTG 65
    |||
Db 238 ACTAG 242

RESULT 14
US-10-198-846-9427/c
; Sequence 9427, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9427
; LENGTH: 822
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 636, 661, 665, 701, 713, 723, 729, 738, 744, 745, 748, 750,
; LOCATION: 762, 777, 778, 785, 792, 794, 817
; OTHER INFORMATION: n = A,T,C or G
; US-10-198-846-9427

Query Match          7.1%; Score 53.2; DB 15; Length 822;
Best Local Similarity 94.8%; Pred. No. 4.1e-06;
Matches 55; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAACCTCGGCCGATATC 58
    |||||
Db 76 ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAACCTCGGCCGCCACC 19
    |||||

RESULT 15
US-10-198-846-6557/c
; Sequence 6557, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
```

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6557
; LENGTH: 863
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 562, 571, 601, 625, 629, 639, 667, 680, 709, 710, 718, 726,
; LOCATION: 746, 749, 758, 762, 766, 792, 794, 798, 803, 809, 813, 816,
; LOCATION: 823, 824, 829, 833, 837, 849, 850, 852, 854
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-6557

Query Match      7.1%; Score 53.2; DB 15; Length 863;
Best Local Similarity 94.8%; Pred. No. 4.2e-06;
Matches 55; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 ATGGCGGCTGTGAGTCTGCGGCTCGGCGACTTGGTGTGGGGGAAACTCGGCCGATATC 58
      |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78 ATGGCGGCTGTGAGTCTGCGGCTCGGCGACTTGGTGTGGGGGAAACTCGGCCGCCACC 21
```

Search completed: December 8, 2003, 15:02:31
Job time : 304 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 8, 2003, 12:48:26 ; Search time 2082 Seconds
(without alignments)
8720.194 Million cell updates/sec

Title: US-09-987-755-1_COPY_34_780
Perfect score: 747
Sequence: 1 atggcggctgtgagctgcg.....aagacctctcttcctagg 747

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	730.2	97.8	1201	13 BX331741	BX331741 BX331741
2	729.4	97.6	3741	11 BC032855	BC032855 Homo sapi
3	727	97.3	870	13 BQ720854	BQ720854 AGENCOURT
4	726.4	97.2	1039	12 BM808810	BM808810 AGENCOURT

5	726	97.2	884	13 BUI58924	BUI58924 AGENCOURT
6	718	96.1	1003	9 AL536040	AL536040 AL536040
7	700.4	93.8	1005	12 BM808969	BM808969 AGENCOURT
8	692.8	92.7	755	12 BI551885	BI551885 603192793
9	673	90.1	778	12 BI600148	BI600148 603246545
10	668.8	89.5	816	12 BI459800	BI459800 603200852
11	647	86.6	1138	12 BM906992	BM906992 AGENCOURT
12	634.2	84.9	993	13 BU520128	BU520128 AGENCOURT
13	633.4	84.8	899	13 BQ892009	BQ892009 AGENCOURT
14	633.2	84.8	890	13 BQ918182	BQ918182 AGENCOURT
15	630.2	84.4	3239	11 AK014456	AK014456 Mus muscu
16	629	84.2	815	12 BI767890	BI767890 603061025
17	623.6	83.5	829	13 BU593961	BU593961 AGENCOURT
18	623	83.4	3250	11 AK040841	AK040841 Mus muscu
19	622.2	83.3	929	13 BQ934274	BQ934274 AGENCOURT
20	614.2	82.2	705	13 BU708303	BU708303 UI-M-FC0-
21	614.2	82.2	813	12 BI691175	BI691175 603314566
22	591.4	79.2	704	13 BQ768940	BQ768940 UI-M-FC0-
23	589.2	78.9	998	10 BB611224	BB611224 BB611224
24	584.4	78.2	773	12 BI103792	BI103792 602887976
25	580.2	77.7	685	10 BB617747	BB617747 BB617747
26	575.2	77.0	682	14 BY735423	BY735423 BY735423
27	567.2	75.9	823	12 BI150279	BI150279 602915496
28	561.2	75.1	801	14 CA512495	CA512495 UI-R-FJ0-
29	553.6	74.1	660	14 BY753847	BY753847 BY753847
30	550.2	73.7	659	12 BM947061	BM947061 UI-M-EH0p
31	547.4	73.3	646	14 BY745429	BY745429 BY745429
32	544.8	72.9	671	12 BI732831	BI732831 603353939
33	544.2	72.9	896	10 BE538528	BE538528 601068185
34	540.4	72.3	732	12 BI218947	BI218947 602938548
35	518.2	69.4	786	12 BG976825	BG976825 602845727
36	517.6	69.3	908	10 BG294400	BG294400 602391505
37	515.8	69.0	775	14 CB601170	CB601170 AGENCOURT
38	506.8	67.8	515	10 BE207054	BE207054 ba09d10.y
39	503.8	67.4	532	9 AL702600	AL702600 DKFZp6861
40	498.4	66.7	642	14 BY745902	BY745902 BY745902
41	497.2	66.6	595	14 CB584042	CB584042 AMGNNUC:N
42	494.4	66.2	816	13 BU373180	BU373180 603589670
43	491.6	65.8	821	12 BG870714	BG870714 602792108
44	490	65.6	574	12 BM288574	BM288574 529889 MA
45	482.8	64.6	632	14 BY743220	BY743220 BY743220

ALIGNMENTS

RESULT 1
BX331741
LOCUS
DEFINITION
BX331741 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
cDNA clone CS0DC001YL09 5-PRIME, mRNA sequence.
ACCESSION
BX331741
VERSION
BX331741.1 GI:30335078
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
1 (bases 1 to 1201)
AUTHORS
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7982.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DC001CF05Q01&cluster=7982.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CS0DC001CF05QP1.	
FEATURES	source
Location/Qualifiers	
1..1201	
/organism="Homo sapiens"	
/mol_type="mRNA"	
/db_xref="taxon:9606"	
/clone="CS0DC001YL09"	
/tissue type="NEUROBLASTOMA COT 25-NORMALIZED"	
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"	
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."	
BASE COUNT	349 a 261 c 332 g 240 t 19 others
ORIGIN	
Query Match 97.8%; Score 730.2; DB 13; Length 1201;	
Best Local Similarity 99.9%; Pred. No. 1.7e-181;	
Matches 730; Conservative 1; Mismatches 0; Indels 0; Gaps 0;	
Qy	1 ATGGCGGCTGTGAGTCTCGGCTCGGCACTTGGTGTGGGGAAACTCGGCCGATATCCT 60
Db	
Qy	75 ATGGCGGCTGTGAGTCTCGGCTCGGCACTTGGTGTGGGGAAACTCGGCCGATATCCT 134
Db	
Qy	61 CCTTGGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAA 120
Db	
Qy	135 CCTTGGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAA 194
Db	
Qy	121 TGCTTCTTTGTGAAATTTTGGACAGAGATCATCGCTGGATCAAGTGAACAGCTG 180
Db	
Qy	195 TGCTTCTTTGTGAAATTTTGGACAGAGATCATCGCTGGATCAAGTGAACAGCTG 254
Db	
Qy	181 AAGCCATATCATGCTCATAAAGAGGAAATGATAAAATTAACAAGGTAAACGATTCCAG 240
Db	
Qy	255 AAGCCATATCATGCTCATAAAGAGGAAATGATAAAATTAACAAGGTAAACGATTCCAG 314
Db	
Qy	241 CAAGCGGTAGATGCTGTGGAAGAGTTCTCAGGAGAGCCAAAGGGAAGACCAAGACGTC 300
Db	
Qy	315 CAAGCGGTAGATGCTGTGGAAGAGTTCTCAGGAGAGCCAAAGGGAAGACCAAGACGTC 374
Db	
Qy	301 TCCACAAATCTTCTGATGACAAGAAATCGACGTAATTCAGTGAGGAGAGAAGTAGGCCA 360
Db	
Qy	375 TCCACAAATCTTCTGATGACAAGAAATCGACGTAATTCAGTGAGGAGAGAAGTAGGCCA 434
Db	
Qy	361 AACTCAGGTGATGAGAGCGCAAACTTAGCCTGTCTGAAGGGGAAGGTGAAGAAACAATG 420
Db	
Qy	435 AACTCAGGTGATGAGAGCGCAAACTTAGCCTGTCTGAAGGGGAAGGTGAAGAAACAATG 494
Db	
Qy	421 GGAGAAAGGAAAGAGAGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTG 480
Db	
Qy	495 GGAGAAAGGAAAGAGAGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTG 554
Db	
Qy	481 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCGGCCCCCAAGGATGAGAAGGAT 540
Db	
Qy	555 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCGGCCCCCAAGGATGAGAAGGAT 614
Db	
Qy	541 CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCTTT 600
Db	
Qy	615 CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCTTT 674
Db	
Qy	601 AAATGGCAGCCCAACCGCAAGCGAGCCTGTTAAAGATGCAGATCCTCATTTCCATTTTC 660
Db	
Qy	675 AAATGGCAGCCCAACCGCAAGCGAGCCTGTTAAAGATGCAGATCCTCATTTCCATTTTC 734
Db	
Qy	661 CTGCTAAGCCCAACAGAGAAGCCAGCTGTCTGTTACCAGGCAATCACGAAGAAGTTGAAA 720
Db	
Qy	735 CTGCTAAGCCCAACAGAGAAGCCAGCTGTCTGTTACCAGGCAATCACGAAGAAGTTGAAA 794
Db	
Qy	721 ATATGTGAAGA 731
Db	
Qy	795 ATATGTGAAGA 805
Db	

BC032855	3741 bp	mRNA	linear	HTC 04-MAR-2003						
Homo sapiens, Similar to RIKEN cDNA 3930401K13 gene, clone										
IMAGE:5265561, mRNA.										
ACCESSION	BC032855									
VERSION	BC032855.1	GI:23958784								
KEYWORDS	HTC.									
SOURCE	Homo sapiens (human)									
ORGANISM	Homo sapiens									
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.									
REFERENCE	1 (bases 1 to 3741)									
AUTHORS	Strausberg,R.									
TITLE	Direct Submission									
JOURNAL	Submitted (07-JUN-2002) National Institutes of Health, Mammalian									
	Gene Collection (MGC), Cancer Genomics Office, National Cancer									
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,									
	USA									
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov									
COMMENT	Contact: MGC help desk									
	Email: cgapbs-r@mail.nih.gov									
	Tissue Procurement: Miklos Palkovits, M.D., Ph.D.									
	cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki									
	Toshiyuki and Piero Carninci (RIKEN)									
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)									
	DNA Sequencing by: Institute for Systems Biology									
	http://www.systemsbio.org									
	contact: amadan@systemsbiology.org									
	Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha									
	Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting									
Clone distribution: MGC clone distribution information can be found										
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov										
Series: IRAK Plate: 47 Row: j Column: 23										
This clone was selected for full length sequencing because it										
passed the following selection criteria: Hexamer frequency ORF										
analysis, Similarity but not identity to protein										
This clone has the following problem: retained intron.										
Location/Qualifiers										
FEATURES	1..3741									
source	/organism="Homo sapiens"									
	/mol_type="mRNA"									
	/db_xref="taxon:9606"									
	/clone="IMAGE:5265561"									
	/tissue type="Testis"									
	/clone_lib="NIH MGC_97"									
	/lab_host="DH10B"									
	/note="Vector: pBluescript"									
BASE COUNT	940 a	959 c	993 g	849 t						
ORIGIN										
Query Match 97.6%; Score 729.4; DB 11; Length 3741;										
Best Local Similarity 99.9%; Pred. No. 3.7e-181;										
Matches 730; Conservative 0; Mismatches 1; Indels 0; Gaps 0;										
Qy	1	ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCCGATATCCT	60							
Db	29									
Qy	61	CCTTGGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAA	120							
Db	89									
Qy	121	TGCTTCTTTGTGAAATTTTGGACAGAGATCATGCTGGATCAAAAGTGAACAGCTG	180							
Db	149									
Qy	181	AAGCCATATCATGCTCATAAAGAGGAAATGATAAAATTAACAAGGTAAACGATTCCAG	240							
Db	209									
Qy	241	CAAGCGGTAGATGCTGTGGAAGAGTTCTCTCAGGAGAGGCTCCAAATCCCCTCTG	300							

```
|||||
Db 269 CAAGCGGTAGTCTGTCGAAGAGTTCTCTCAGGAGAGCCAAAGGAAAGACACAGACGTCA 328
QY 301 TCCCAACAATTCCTCTGATGACAAAGATCGACGTAATCCAGTCAGGAGAGAGTAGGCCA 360
Db 329 TCCCAACAATTCCTCTGATGACAAAGATCGACGTAATCCAGTCAGGAGAGAGTAGGCCA 388
QY 361 AACTCAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGTAAGGGAAGGTGAAGAAACATG 420
Db 389 AACTCAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGTAAGGGAAGGTGAAGAAACATG 448
QY 421 GGAGAAGGAAGAGAGGGTGTCTTTCAGGCTCTTTCAGAGAGAGGCTCCAAATCCCCTCTG 480
Db 449 GGAGAAGGAAGAGAGGGTGTCTTTCAGGCTCTTTCAGAGAGAGGCTCCAAATCCCCTCTG 508
QY 481 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGGTGCGCCCAAGGATGAGAAGGAT 540
Db 509 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGGTGCGCCCAAGGATGAGAAGGAT 568
QY 541 CTCACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCCGATGGCCGCGTTT 600
Db 569 CTCACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCCGATGGCCGCGTTT 628
QY 601 AAATGGCAGCCCAAGCGAGCGCTGTTAAAGATGCAGATCCTCATTTCCATCATTTTC 660
Db 629 AAATGGCAGCCCAAGCGAGCGCTGTTAAAGATGCAGATCCTCATTTCCATCATTTTC 688
QY 661 CTGCTAAGCCCAACAGAGAAGCCAGCTGTCTGTACAGGCAATCAAGAAAGTTGAAA 720
Db 689 CTGCTAAGCCCAACAGAGAAGCCAGCTGTCTGTACAGGCAATCAAGAAAGTTGAAA 748
QY 721 ATATGTGAAGA 731
Db 749 ATATGTGAAGA 759
```

RESULT 3

BO720854
LOCUS BO720854 870 bp mRNA linear EST 16-JUL-2002
DEFINITION AGENCOURT 8241262 Lupski_sympathetic_trunk Homo sapiens cDNA clone
IMAGE:6187142 5', mRNA sequence.

ACCESSION BO720854.1 GI:21859751

VERSION BO720854.1

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE NIH-MGC http://mgi.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. James R. Lupski

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM13580 row: k column: 15

High quality sequence stop: 668.

Location/Qualifiers

1. .870

source

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6187142"

/sex="male"

/tissue_type="sympathetic trunk"

/dev_stage="adult, 16 yr"

/lab_host="DH10B"

RESULT 4

BM808610

/clone lib="Lupski_sympathetic_trunk"
/note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site 2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGCGTCCG-3' and
5'-GACTAGTTCTAGATCGGAGCGGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.9 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine); available through Life
Technologies."

BASE COUNT 270 a 193 c 245 g 161 t 1 others

ORIGIN

Query Match 97.3%; Score 727; DB 13; Length 870;

Best Local Similarity 100.0%; Pred. No. 1.le-180;

Matches 727; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CGGCTGTGAGTCTCGGCTCGCGACTTGGTGTGGGGAAACTCGGCCGATATCCTCCTT 64

Db 1 CGGCTGTGAGTCTCGGCTCGCGACTTGGTGTGGGGAAACTCGGCCGATATCCTCCTT 60

QY 65 GGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAATGCT 124

Db 61 GGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAATGCT 120

QY 125 TCTTTGTGAAATTTTGGAAACAGAAAGATCATGCTTGAAGTGAACAGCTGAAGC 184

Db 121 TCTTTGTGAAATTTTGGAAACAGAAAGATCATGCTTGAAGTGAACAGCTGAAGC 180

QY 185 CATATCATGCTCATAAAGAGGAAATGATAAAATTAACAAGGTAAACGATTCCAGCAAG 244

Db 181 CATATCATGCTCATAAAGAGGAAATGATAAAATTAACAAGGTAAACGATTCCAGCAAG 240

QY 245 CGGTAGATGCTGTCGAAGAGTTCTCTCAGGAGAGCCAAAGGAAAGACCAGACGTCATCCC 304

Db 241 CGGTAGATGCTGTCGAAGAGTTCTCTCAGGAGAGCCAAAGGAAAGACCAGACGTCATCCC 300

QY 305 ACAATTTCTTCTGATGACAAGATCGACGTAATTCAGTGAGGAGAGAAAGTAGGCCAAACT 364

Db 301 ACAATTTCTTCTGATGACAAGATCGACGTAATTCAGTGAGGAGAGAAAGTAGGCCAAACT 360

QY 365 CAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAACATGGGAG 424

Db 361 CAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAACATGGGAG 420

QY 425 AAGGAAAGAGAGGGTGTCTTCAGGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTGAAA 484

Db 421 AAGGAAAGAGAGGGTGTCTTCAGGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTGAAA 480

QY 485 GAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGAGGATGAGAGGATCTCA 544

Db 481 GAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGAGGATGAGAGGATCTCA 540

QY 545 CCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGGACCGATGGCCCGGTTAAAT 604

Db 541 CCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGGACCGATGGCCCGGTTAAAT 600

QY 605 GGCAGCCAAACCCGAAAGCGAGCTGTTAAAGATGCAGATCCTCATTTCCATCATTTCTGC 664

Db 601 GGCAGCCAAACCCGAAAGCGAGCTGTTAAAGATGCAGATCCTCATTTCCATCATTTCTGC 660

QY 665 TAAGCCAAACAGAGAAGCCAGCTGTCTGTTACAGGCAATCACGAAGAGTTGAAAATAT 724

Db 661 TAAGCCAAACAGAGAAGCCAGCTGTCTGTTACAGGCAATCACGAAGAGTTGAAAATAT 720

QY 725 GTGAAGA 731

Db 721 GTGAAGA 727

LOCUS BM808810 1039 bp mRNA linear EST 05-MAR-2002
DEFINITION AGENCOURT 6617726 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5734324
5', mRNA sequence.
ACCESSION BM808810
VERSION BM808810.1 GI:19125633
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1039)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Invitrogen
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12739 row: h column: 05
High quality sequence stop: 715.
Location/Qualifiers
1..1039
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5734324"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_124"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: EcoRV
(destroyed); Site_2: NotI; RNA source male hippocampus,
age 27. Library is oligo-dT primed and directionally
cloned (EcoRV site is destroyed upon cloning). Average
insert size 1.4 kb, insert size range 0.9-4 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 012."
BASE COUNT 307 a 247 c 288 g 197 t
ORIGIN
Query Match 97.2%; Score 726.4; DB 12; Length 1039;
Best Local Similarity 99.9%; Pred. No. 1.6e-180;
Matches 727; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 4 GCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCGGATATCCTCT 63
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
9 GCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCGGATATCCTCT 68
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 64 TGGCCAGGAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAGAAATGC 123
69 TGGCCAGGAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAGAAATGC 128
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 124 TTCTTTTGAAATTTTGTGAAACAGAGATCATGCCTGGATCAAAGTGAACAGCTGAAG 183
129 TTCTTTTGAAATTTTGTGAAACAGAGATCATGCCTGGATCAAAGTGAACAGCTGAAG 188
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 184 CCATATCATGCTCATAAAGAGGAATGATAAAATTAACAAGGGTAACGATTCAGCAA 243
189 CCATATCATGCTCATAAAGAGGAATGATAAAATTAACAAGGGTAACGATTCAGCAA 248
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 244 GCGGTAGTGTGCGAGAGTTCTCAGGAGAGCCAAAGGGAAAGACCAGACGTATCC 303
249 GCGGTAGTGTGCGAGAGTTCTCAGGAGAGCCAAAGGGAAAGACCAGACGTATCC 308
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
304 CACAATTTCTGATGACAAGAATCGACGTAATTCAGTGAGGAGAGAGTAGGCCAAAC 363
309 CACAATTTCTGATGACAAGAATCGACGTAATTCAGTGAGGAGAGAGTAGGCCAAAC 368

Qy 364 TCAGGTGATGAGAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAACATGGGA 423
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
369 TCAGGTGATGAGAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAACATGGGA 428
Qy 424 GAAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTGAAA 483
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
429 GAAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTGAAA 488
Qy 484 AGAGCCCAAGAGCAAAAGTCCCCGGAAGCGGGTCCGCCCCCAAGAGGATGAGAAGGATCTC 543
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
489 AGAGCCCAAGAGCAAAAGTCCCCGGAAGCGGGTCCGCCCCCAAGAGGATGAGAAGGATCTC 548
Qy 544 ACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGCGCGGACCGCGGCTTTAAA 603
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
549 ACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGCGCGGACCGCGGCTTTAAA 608
Qy 604 TGGCAGCCCAACCGCAAGCGAGCCTGTAAAGATGAGATCCTCATTTCCATCTTCCTG 663
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
609 TGGCAGCCCAACCGCAAGCGAGCCTGTAAAGATGAGATCCTCATTTCCATCTTCCTG 668
Qy 664 CTAAGCCCAACAGAGAGCCAGCTGTCTTACCAGCAATCAGCAAGAAAGTTGAAATA 723
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
669 CTAAGCCCAACAGAGAGCCAGCTGTCTTACCAGCAATCAGCAAGAAAGTTGAAATA 728
Qy 724 TGTGAAGA 731
Db ||||||||
729 TGTGAAGA 736
RESULT 5
BU158924
LOCUS
DEFINITION AGENCOURT 7844927 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6044084
5', mRNA sequence.
ACCESSION BU158924
VERSION BU158924.1 GI:22672834
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 884)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13286 row: b column: 21
High quality sequence stop: 776.
Location/Qualifiers
1..884
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6044084"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_92"
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 274 a 195 c 244 g 169 t 2 others
ORIGIN

Qy 600 TAAATGGCAGCCAAACCGCAAGCGAGCCCTGTTAAAGATGCAGATCCTCAATTTCCATCATTT 659
Db 664 TAAATGGCAGCCAAACCGCAAGCGAGCCCTGTTAAAGATGCAGATCCTCAATTTCCATCATTT 723
Qy 660 CCTGCTAAGCCAAACAGAGAAGCCAGCTGTCTGTTACAGGCAATCAGAAAGATTGAA 719
Db 724 CCTGCTAAGCCAAACAGAGAAGCCAGCTGTCTGTTACAGGCAATCAGAAAGATTGAA 783
Qy 720 AATATGTGAAGA 731
Db 784 AATATGTGAAGA 795

RESULT 7
BM808969
LOCUS
DEFINITION BM808969 1005 bp mRNA linear EST 05-MAR-2002
AGENCOURT 6617730 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5734420
5', mRNA sequence.
ACCESSION BM808969
VERSION BM808969.1 GI:19125792
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1005)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Invitrogen
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM12739 row: 1 column: 05
High quality sequence stop: 639.
Location/Qualifiers
1..1005
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5734420"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH MGC_124"
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: EcoRV
(destroyed); Site 2: NotI; RNA source male hippocampus,
age 27. Library is oligo-dT primed and directionally
cloned (EcoRV site is destroyed upon cloning). Average
insert size 1.4 kb, insert size range 0.9-4 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 012."

BASE COUNT 299 a 235 c 278 g 193 t
ORIGIN

Query Match 93.8%; Score 700.4; DB 12; Length 1005;
Best Local Similarity 98.9%; Pred. No. 1.2e-173;
Matches 716; Conservative 0; Mismatches 6; Indels 2; Gaps 1;

Qy 1 ATGGCGGCTGTGAGTCTGCGGCTCGGCGACTTGGTGTGGGGAACTCGGCCGATATCCT 60
Db 20 ATGGCGGCTGTGAGTCTGCGGCTCGGCGACTTGGTGTGGGGAACTCGGCCGATATCCT 79
Qy 61 CCTTGGCCAGGAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAGAA 120
Db 80 CCTTGGCCAGGAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAGAA 139

Qy 121 TGCTTCTTTGTGAATTTTGGACAGAAAGATCATGCTCGCTGGATCAAAGTGAACAGCTG 180
Db 140 TGCITCTTTGTGAATTTTGGACAGAAAGATCATGCTCGCTGGATCAAAGTGAACAGCTG 199
Qy 181 AAGCCATATCATGCTCATAAAGAGGAAATGATAAAATTAACAAGGGTAAACGATTCCAG 240
Db 200 AAGCCATATCATGCTCATAAAGAGGAAATGATAAAATTAACAAGGGTAAACGATTCCAG 259
Qy 241 CAAGCGGTAGATGCTGTGGAAGAGTTCCTCAGGAGAGCCAAAGGGAAAGACCAGACGTCA 300
Db 260 CAAGCGGTAGATGCTGTGGAAGAGTTCCTCAGGAGAGCCAAAGGGAAAGACCAGACGTCA 319
Qy 301 TCCACAAATTTCTTCTGATGACAAGAAATCGACGTAATTCAGTGAGGAGAGAGTAGGCCA 360
Db 320 TCCACAAATTTCTTCTGATGACAAGAAATCGACGTAATTCAGTGAGGAGAGAGTAGGCCA 379
Qy 361 AACTCAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGAAGGGAGGTGAAGAAGACATG 420
Db 380 AACTCAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGAAGGGAGGTGAAGAAGACATG 439
Qy 421 GGAGAAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTG 480
Db 440 GGAGAAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTG 499
Qy 481 AAAGAGCCCCAAGAGCAAAAGTCCCGGAAGCGGGGTGGCCCCCAAAGGATGAGAAGGAT 540
Db 500 AAAGAGCCCCAAGAGCAAAAGTCCCGGAAGCGGGGTGGCCCCCAAAGGATGAGAAGGAT 559
Qy 541 CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCGTTT 600
Db 560 CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCGTTT 619
Qy 601 AATGGCAGCCAAACCGCAAGCGAGCTGTTAAAGATGCAGATCCTCATTTCCATCATTT 660
Db 620 AATGGCAGCCAAACCGCAAGCGAGCTGTTAAAGATGCAGATCCTCATTTCCATCATTT 679
Qy 661 CTGCTAAGCCAAACAGAGAAGCCAGCTGTCTGTACCAGGC--AATCAGGAAGAAGTTGA 718
Db 680 CTGCTAAGCCAAACAGAGAAGCCAGCTGTCTGTACCAGGCCAATCACCGAAGAAGTTGA 739
Qy 719 AAT 722
Db 740 AAT 743

RESULT 8
BI551885
LOCUS
DEFINITION BI551885 755 bp mRNA linear EST 05-SEP-2001
603192793F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5263991 5',
mRNA sequence.
ACCESSION BI551885
VERSION BI551885.1 GI:15439197
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 755)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11665 row: j column: 24
High quality sequence stop: 746.

FEATURES
source

Location/Qualifiers
 1. 755
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5263991"
 /tissue_type="hippocampus"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_95"
 /notes="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgaag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
 236 a 156 c 218 g 145 t

BASE COUNT
ORIGIN

Query Match 92.7%; Score 692.8; DB 12; Length 755;
 Best Local Similarity 99.3%; Pred. No. 1.1e-171;
 Matches 727; Conservative 0; Mismatches 2; Indels 3; Gaps 3;

QY	1	ATGCGGGCTGTGAGTCTCGGGCTCGCGGACCTTGGTGTGGGGAAAACTCGGCCGATATCCT	60
DB	25	ATGCGGGCTGTGAGTCTCGGGCTCGCGGACCTTGGTGTGGGGAAAACTCGGCCGATATCCT	84
QY	61	CCTTGGCCAGGAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAAA	120
DB	85	CCTTGGCCAGGAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAAA	144
QY	121	TGCTTCTTTGTGAAATTTTTGGAAACAGAAAGATCATGCCCTGGATCAAAAGTGGAAACAGCTG	180
DB	145	TGCTTCTTTGTGAAA-TTTTTGGAAACAGAAAGATCATGCCCTGGATCAAAAGTGGAAACAGCTG	203
QY	181	AAGCCATATCATGCTCATAAAGAGGAAATGATAAAAAATTAAACAAGGGTAAACGATTCCAG	240
DB	204	AAGCCATATCATGCTCATAAAGAGGAAATGATAAAAAATTAAACAAGGGTAAACGATTCCAG	263
QY	241	CAAGCGGTAGATGCTGTGGAAGAGTTCTCAGGAGAGCCAAAGGAAAAAGACCAAGACGTCA	300
DB	264	CAAGCGGTAGATGCTGTGGAAGAGTTCTCAGGAGAGCCAAAGGAAAAAGACCAAGACGTCA	323
QY	301	TCCACAATTTCTTGATGACAAGAAATCGACGTAATTCCAGTGAGGAGAGAGAGTAGGCCA	360
DB	324	TCCACAATTTCTTGATGACAAGAAATCGACGTAATTCCAGTGAGGAGAGAGAGTAGGCCA	383
QY	361	AACTCAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAGAACATG	420
DB	384	AACTCAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAGAACATG	443
QY	421	GGAGAAGGAAAGAGAGGGGTGCTTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG	480
DB	444	GGAGAAGGAAAGAGAGGGGTGCTTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG	503
QY	481	AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGGTGGCCCCCAAAAGGATGAGAAGGAT	540
DB	504	AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGGTGGCCCCCAAAAGGATGAGAAGGAT	563
QY	541	CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCGTTT	600
DB	564	CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCGTTT	623
QY	601	AAATGSCAGCCAAACCGCAAGCGAGCCTGTTAAAGATGCAGATCCTCATTTCCATCA-TTT	659
DB	624	AAATGSCAGCCAAACCGCAAGCGAGCCTGTTAAAGATGCAGATCCTCATTTCCATCA-TTT	683
QY	660	CCTGCTAAGCC-AAACAGAGAAGCCAGCTGTCTGTTACCAGGCAATCACGAAGAAGTTGA	718
DB	684	CCTGCTAAGCCAAACAGAGAAGCCAGCTGTCTGTTACCAGGCAATCACGAAGAAGTTGA	743

Qy	719	AAATATGTGAAG	730
Db	744	AAATATGTGAAG	755
RESULT 9			
BI600148			
LOCUS			
DEFINITION	BI600148	603246545F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5288915 5', mRNA linear EST 07-SEP-2001 mRNA sequence.	
ACCESSION	BI600148	GI:15493087	
VERSION	EST.		
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 778) NIH-MGC http://mgc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: Miklos Palkovits, M.D., Ph.D. CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN) CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM11730 row: i column: 12 High quality sequence stop: 773. Location/Qualifiers 1..778 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:5288915" /tissue_type="hypothalamus" /lab_host="DH10B" /clone_lib="NIH_MGC_96" /note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag); Oligo-dr primed using primer 5'-TTT-TTTTTTTTTTTT-VN-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."		
BASE COUNT	237 a	162 c	227 g
ORIGIN			
Query Match	90.1%	Score 673;	DB 12; Length 778;
Best Local Similarity	98.2%	Pred. No. 1.8e-166;	
Matches 712;	Conservative 0;	Mismatches 10;	Indels 3; Gaps 3;
Qy	1	ATGGCGGCTGTGAGTCTCGGCTCGGACTTGTTGTTGGGAAAACCTCGGCCGATATCCT	60
Db	26	ATGGCGGCTGTGAGTCTCGGCTCGGACTTGTTGTTGGGAAAACCTCGGCCGATATCCT	85
Qy	61	CCTTGGCCAGGAAAGATTGTTAATCCACCAGGACTTGAAGAAACCTCGCGAAAGAAA	120
Db	86	CCTTGGCCAGGAAAGATTGTTAATCCACCAGGACTTGAAGAAACCTCGCGAAAGAAA	145
Qy	121	TGCTTCCTTTGTGAAATTTTTTGGAACAGAGATCATGCGCTGGATCAAAGTGGAACAGCTG	180
Db	146	TGCTTCCTTTGTGAAATTTTTTGGAACAGAGATCATGCGCTGGATCAAAGTGGAACAGCTG	205
Qy	181	AAGCCATATCATGCTCATAAGAGGAAATGATAAAAAATTAAACAAGGGTAACGATTCCAG	240
Db	206	AAGCCATATCATGCTCATAAGAGGAAATGATAAAAAATTAAACAAGGGTAACGATTCCAG	265

QY 241 CAAGCGGTAGATGCTGTGGAAGAGTTCTCTCAGGAGAGCCAAAGGGAAGACACGACGTCA 300
|||||
Db 266 CAAGCGGTAGATGCTGTGGAAGAGTTCTCTCAGGAGAGCCAAAGGGAAGACACGACGTCA 325
|||||
QY 301 TCCACAAATCTTCTGATGACAAGAAATCGACGTAATTCAGTGAAGGAGAGAGTAGGCCA 360
|||||
Db 326 TCCACAAATCTTCTGATGACAAGAAATCGACGTAATTCAGTGAAGGAGAGAGTAGGCCA 385
|||||
QY 361 AACTCAGGTGATGAGAAAGCGCAAACTTAGCTGTCTGAAGGGAAGGTGAAGAACAATG 420
|||||
Db 386 AACTCAGGTGATGAGAAAGCGCAAACTTAGCTGTCTGAAGGGAAGGTGAAGAACAATG 445
|||||
QY 421 GGAGAAGGAAAGAGGGTGTCTTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTG 480
|||||
Db 446 GGAGAAGGAAAGAGGGTGTCTTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTGTG 505
|||||
QY 481 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGAGATGAGAAGAT 540
|||||
Db 506 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCG-CCCCAAAGATGAGAAGAT 564
|||||
QY 541 CTCACCATCCGGAGTCTA-GTACCGTGAAGGGGATGATGGCCGGAACCGATGGCCGCTT 599
|||||
Db 565 CTCACCATCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGGAACCGATGGCCGCTT 624
|||||
QY 600 TAAATGGAGCCCAACCGCAAGCGAGCCCTGTAAAGATGAGATCCTCAATTCATCATTT 659
|||||
Db 625 TAAATGGAGCCCAACCGCAAGCGAGCCCTGTAAAGATGAGATCCTCAATTCATCATTT 684
|||||
QY 660 CCTGCTAAGCCAAACAGAGAAGCCAGCTGTCTGTTACCAAGCAATCACGAAGAAGTTGAA 719
|||||
Db 685 CCTGCTAAGCCAAACAGAGAAGCCAGCTGTCTGTTACCAAGC-ATCACGAAGAAGTTGAA 743
|||||
QY 720 AATAT 724
|||||
Db 744 AATAT 748

RESULT 10
BI459800
LOCUS 603200852F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5266620 5',
EST.
ACCESSION BI459800
VERSION BI459800.1 GI:15250456
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 816)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM1672 row: h column: 13
High quality sequence stop: 719.
Location/Qualifiers
1..816
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5266620"
/lab_host="DH10B"

/clone lib="NIH_MGC_97"
/note="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 254 a 181 c 230 g 151 t
ORIGIN

Query Match 89.5%; Score 668.8; DB 12; Length 816;
Best Local Similarity 97.9%; Pred. No. 2.4e-165;
Matches 709; Conservative 0; Mismatches 12; Indels 3; Gaps 3;
QY 1 ATGGCGGCTGTAGTCTGGGCTCGGCGACTTGGTGTGGGGGAAACTCGGCCGATATCCT 60
|||||
Db 9 ATGGCGGCTGTAGTCTGGGCTCGGCGACTTGGTGTGGGGGAAACTCGGCCGATATCCT 68
|||||
QY 61 CCTTGGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAA 120
|||||
Db 69 CCTTGGCCAGGAAAGA-TGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAA 127
|||||
QY 121 TGCTTCTTTGTGAAATTTTGGAAACAGAAAGATCATGCCTGGATCAAAGTGGAACAGCTG 180
|||||
Db 128 TGCTTCTTTGTGAAATTTTGGAAACAGAAAGATCATGCCTGGATCAAAGTGGAACAGCTG 187
|||||
QY 181 AAGCCATATCATGCTCATAAAGAGGAAATGATAAAATTAACAAGGTAAACGATTCCAG 240
|||||
Db 188 AAGCCATATCATGCTCATAAAGAGGAAATGATAAAATTAACAAGGTAAACGATTCCAG 247
|||||
QY 241 CAAGCGGTAGATGCTGTGGAAGAGTTCTCTCAGGAGAGCCAAAGGGAAGACACGACGTCA 300
|||||
Db 248 CAAGCGGTAGATGCTGTGGAAGAGTTCTCTCAGGAGAGCCAAAGGGAAGACACGACGTCA 307
|||||
QY 301 TCCACAAATCTTCTGATGACAAGAAATCGACGTAATTCAGTGAAGAGAGAAAGTAGGCCA 360
|||||
Db 308 TCCACAAATCTTCTGATGACAAGAAATCGACGTAATTCAGTGAAGAGAGAAAGTAGGCCA 367
|||||
QY 361 AACTCAGGTGATGAGAAGCGCAAACTTAGCTGTCTGAAGGGAAGGTGAAGAACAATG 420
|||||
Db 368 AACTCAGGTGATGAGAAGCGCAAACTTAGCTGTCTGAAGGGAAGGTGAAGAACAATG 427
|||||
QY 421 GGAGAAGGAAAGAGAGGGTGTCTTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTG 480
|||||
Db 428 GGAGAAGGACAGAGAGGGTGTCTTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTG 487
|||||
QY 481 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGAGATGAGAAGAT 540
|||||
Db 488 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCG-CCCCAAAGAGATGAGAAGAT 546
|||||
QY 541 CTCACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGGCCGGAACCGATGGCCGCTT 600
|||||
Db 547 CTCACCGTCCCGAGTCTAGTACCGTGAAGGGGATGATGGCCGGAACCGATGGCCGCTT 606
|||||
QY 601 AAATGGCAGCAACCGCAAGCGAGCTGTTAAAGATGCAGATCCTCATTTCCATCATTTTC 660
|||||
Db 607 AAATGGCAG-CAACGCAAGCGAGCTGTTAAAGATGGCGATCCTCATTTCCATCATTTTC 665
|||||
QY 661 CTGCTAAGCCAAACAGAGAAGCCAGCTGTCTGTTACAGGCAATCACGAAGAAGTTGAAA 720
|||||
Db 666 CTGCTAAGCAACAGGAGAAGCCAGCTGTCTGTTACAGGCAATCACGAAGAAGTTGACC 725
|||||
QY 721 ATAT 724
|||||
Db 726 ATTT 729

RESULT 11
BM906992
LOCUS

BM906992 1138 bp mRNA linear EST 12-MAR-2002

Db 1 GCGGCTGTGAGTCTGCGGCTCGGCGACTTGGTGTGGGGAACCTGGGCGGTATCCTCCC 60

Qy 64 TGGCCAGGAAGATTGTTAATCCACCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAATGC 123

Db 61 TGGCCAGGAAGATTGTTAATCCACCACCAAGGACTTAAAGAAACCAAGTGGAAAGAAATGC 120

Qy 124 TTCTTTGTGAAATTTTGTGAAACAGAGATCATGCTGGATCAAAAGTGAACAGCTGAAG 183

Db 121 TTCTTTGTGAAATTTTGTGAAACAGAGATCATGCTGGATCAAAAGTGAACAGCTGAAG 180

Qy 184 CCATATCATGCTCATAAAGAGGAATGATAAAATTAACAAGGGTAAACGATTCCAGCAA 243

Db 181 CCTTACCATGCTCACAGGAGGAGATGATAAGATTAAACAAGGGTAAACGGTTCCAGCAA 240

Qy 244 GCGGTAGATGCTGTGCAAGAGTTCTTCAGGAGAGCCAAAGGAAAGACAGACGTCTATCC 303

Db 241 GCTGTGGATGCTGTTGAAGAGTTCTTCAGGAGAGCCAAAGGAAAGACAGACATCATCC 300

Qy 304 CACAATTCTTCTGATGACAAGAATCGACGTAATTCAGTGAGGAGAGTAGGCCAATAC 363

Db 301 CACACTTCTGCTGATGACAAGAATCGGCGTAATTCAGTGAGGAGAGTAGGCCAATAC 360

Qy 364 TCAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAGAACATGGGA 423

Db 361 TCAGGTGATGAGAAACGCAAGCTTAGCCTGTCTGAAGGGAAGGTGAAGAGAACATGGGA 420

Qy 424 GAAGGAAGAAGAGGTGTCTTCAGGCTTTCAGAGAGAGGCTCCAATCCCTCTGAAA 483

Db 421 GAAGGAAGAAGAGGTGTCTTCAGGCTTTCAGAGAGAGGCTCCAATGC---CTTAAA 477

Qy 484 AGAGCCCAAGAGCAAAGTCCCGGAAGCGGGGTGCGGCCCAAGAGGATGAGAGGATCTC 543

Db 478 AGAGCCCAAGAGCAAAGTCCCGGAAGCGGGGTGCGGCCCAAGAGGATGAGAGGACCTC 537

Qy 544 ACCATCCCGAGTCTAGTACCGTGAAGGGATGATGGCGGACCGATGGCCGCGTTTAAA 603

Db 538 ACCATCCCTGAGTCTAGCACTGTAAAGGGATGATGGCTGGACCGATGGCTGCATTANA 597

Qy 604 TGGCAGCCAAACCGCAAGCGAGCCTGTTAAAGATGCAGATCCTCATTTTCCATCATTTCTG 663

Db 598 TGGCAGCCAAACAGCGACCGAGCCAGTCAAAGATGCAGATCCTCATTTTCCATCATTTCTG 657

Qy 664 CTAAGCCAAACAGAGAGCCAGCTGTCTGTTACCAGGCAATCACGAAGATTGAAATA 723

Db 658 TTGAGCCAAACAGAGAGCCAGCTGTCTGTTACCAGGCAATCACAAAGAGTTGAAATA 717

Qy 724 TGTGAAGA 731

Db 718 TGTGAAGA 725

RESULT 13
BQ892009
LOCUS BQ892009 899 bp mRNA linear EST 16-AUG-2002
DEFINITION AGENCOURT 8754837 NIH_MGC_129 Mus musculus cdna clone IMAGE:6391577
5', mRNA sequence.
ACCESSION BQ892009
VERSION BQ892009.1 GI:22284023
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 899)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
cdna Library Preparation: ResGen, Invitrogen Corp
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation

FEATURES
source
1..899
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6391577"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_129"
/note="Organ: o1factory epithelium; Vector:
pCMV-SPORT6.1.cdb; Site_1: EcoRV; Site_2: NotI; Cloned
unidirectionally. Primer: Oligo dt. Average insert size
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
is a NIH_MGC Library."

BASE COUNT 277 a 190 c 257 g 174 t 1 others
ORIGIN

Query Match 84.8%; Score 633.4; DB 13; Length 899;
Best Local Similarity 92.6%; Pred. No. 5.2e-156;
Matches 677; Conservative 0; Mismatches 51; Indels 3; Gaps 1;

Qy 1 ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAACCTCGGCCGATATCCT 60

Db 5 ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAACCTCGGCCGATATCCT 64

Qy 61 CCTGGCCAGGAAGATTGTTAATCCACCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAA 120

Db 65 CCCTGGCCAGGAAGATTGTTAATCCACCACCAAGGACTTAAAGAAACCAACGCTGGAAGAAA 124

Qy 121 TGCTTCTTTGTGAAATTTTGGACAGAAAGATCATGCTTGATCAAAAGTGAACAGCTG 180

Db 125 TGCTTCTTTGTGAAATTTTGGACAGAAAGATCATGCTTGATCAAAAGTGAACAGCTA 184

Qy 181 AAGCCATATCATGCTCATAAAGAGGAAATGATAAAATTAACAAGGTAAACGATTCCAG 240

Db 185 AAGCCTTACCATGCTCACAAAGGAGAGATGATAAAGATTAAACAAGGTAAACGTTCCAG 244

Qy 241 CAAGCGGTAGATGCTGTGAAAGAGTTCTCAGGAGAGCCAAAGGGAAGAACCCAGACGTCA 300

Db 245 CAAGCTGTGATGCTGTGAAAGAGTTCTCAGGAGAGCCAAAGGGAAGAACCCAGACATCA 304

Qy 301 TCCCACAATTTCTTCTGATGACAAGAATCGACGTAAATCCAGTGAGGAGAGAAGTAGCCA 360

Db 305 TCCCACTTCTGCTGATGACAAGAATCGGCGTAAATCCAGTGAGGAGAGAAGTAGSCCA 364

Qy 361 AACTCAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAGAACATG 420

Db 365 AACTCAGGTGATGAGAAGCGCAAGTTAGCCTGTCTGAAGGGAAGGTGAAGAAGAACATG 424

Qy 421 GGAGAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCTCTG 480

Db 425 GGAGAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATGC---CTT 481

Qy 481 AAAAGAGCCCAAGCAAAAGTCCCGGAAGCGGGGTGCGGCCCCCAAGAGGATGAGAAGGAT 540

Db 482 AAAAGAGCCCAAGCAAAAGTCCCGGAAGCGGGGTGCGGCCCCCAAGAGGATGAGAAGGAG 541

Qy 541 CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGAGCCGATGGCCCGGTTT 600

Db 542 CTCACCATCCCTGAGTGTAGCACTGTAAAGGGGAGGATGGCTGGACCGATGGCTGCAATT 601

Qy 601 AAATGGCAGCCAAACCGCAAGCGAGCCTGTTAAAGATGCAGATCCTCATTTTCCATCATTT 660

Db 602 AAATGGCAGCCAAACCGCAAGCGAGCCTGTTAAAGATGCAGATCCTCATTTTCCATCATTT 661

Qy 661 CTGCTAAGCCAAACAGAGAGCCAGCTGTCTGTACCAGGCAATCACGAAGAGATTGAAA 720

Db 662 CTGTTGAGCCGAACAGAGAGCCAGCTGTCTGTACCAGGCAATCACAAAGAGATTGAAA 721

```
Qy 721 ATATGTGAAGA 731
|||||
Db 722 ATATGTGAAGA 732

RESULT 14
BQ918182
LOCUS BQ918182 890 bp mRNA linear EST 20-AUG-2002
DEFINITION AGENCOURT 8765993 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6330861
5', mRNA sequence.
ACCESSION BQ918182
VERSION BQ918182.1 GI:22332892
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 890)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
Ph.D.
cDNA Library Preparation: ResGen, Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13785 row: o column: 22
High quality sequence stop: 648.
FEATURES
source
1..890
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6330861"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_130"
/note="Organ: otocysts; Vector: pCMV-SPORT6.1.cdb;
Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally.
Primer: Oligo dT. Average insert size 1.95 kb.
Constructed by ResGen, Invitrogen Corp. Note: this is a
NIH MGC Library."
BASE COUNT 268 a 201 c 245 g 175 t 1 others
ORIGIN
Query Match 84.8%; Score 633.2; DB 13; Length 890;
Best Local Similarity 92.8%; Pred. No. 5.9e-156;
Matches 675; Conservative 0; Mismatches 49; Indels 3; Gaps 1;
Qy 5 CGGCTGTGAGTCTGCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCGGATATCCTCTT 64
|||||
Db 1 CGGCTGTGAGTCTGCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCGGATATCCTCTT 60
|||||
Qy 65 GGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAATGCT 124
|||||
Db 61 GGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGGTGAAAGAAATGCT 120
|||||
Qy 125 TCTTTGTGAATTTTTTGGACAGAGATCATGCTGATCAAGTGGACAGCTGAAGC 184
|||||
Db 121 TCTTTGTGAAGTTTTTGGACAGAGATCATGCTGATCAAGTGGACAGCTGAAGC 180
|||||
Qy 185 CATATCATGCTCATAAAGAGGAAATGATAAAATTAACAGGGTAAACGATCCAGCAG 244
|||||
Db 181 CTTACCATGCTCACAAGGAGGAGATGATAAAGATTAAAGGGTAAACGGTTCAGCAG 240
|||||
Qy 245 CGGTAGATGCTGTCGAGAGTTCTCTCAGGAGAGCCAAAGGGAAAGACGACGTCATCC 304
|||||
Db 241 CTGTGGATGCTGTTGAAGAGTTCTCTCAGGAGAGCCAAAGGGAAAGACGACATCCTC 300
|||||
```

```
Qy 305 ACAATTCTTCTGATGACAAGAATCGACGTAATTCCAGTGAGGAGAGAGTAGGCCAAACT 364
|||||
Db 301 ACACCTTCTGCTGATGACAAGAATCGGCGTAATTCCAGTGAGGAGAGAGTAGGCCAAACT 360
|||||
Qy 365 CAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAGAACATGGAG 424
|||||
Db 361 CAGGTGATGAGAAGCGCAAGCTTAGCCTGTCTGAAGGGAAGGTGAAGAAGAACATGGAG 420
|||||
Qy 425 AAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAAAA 484
|||||
Db 421 AAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATGC--CTTAAAA 477
|||||
Qy 485 GAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGAGATGAGAAGGATCTCA 544
|||||
Db 478 GAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGAGATGAGAAGGATCTCA 537
|||||
Qy 545 CCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGCGCGGACCGATGCGCGCTTTAAAT 604
|||||
Db 538 CCATCCCTGAGTCTAGCACTGTAAAGGGGATGATGCTGGACCGATGCTCATTTCTGT 597
|||||
Qy 605 GGCAAGCAACCGCAAGCGAGCCTGTAAAGATGCAGATCCTCATTTCCATCATTTCTGTC 664
|||||
Db 598 GGCAGCAACAGCGACCGAGCCAGTCAAAAGATGCAGATCCTCATTTCTCATTTCTGT 657
|||||
Qy 665 TAAGCCAAACAGAGAGCCAGCTGTCTGTACCGCAATCACGAAGAAGTTGAAAAATAT 724
|||||
Db 658 TGAGCCAAACAGAGAGCCAGCTGTCTGTACCGCAATCACGAAGAAGTTGAAAAATAT 717
|||||
Qy 725 GTGAAGA 731
|||||
Db 718 GTGAAGA 724

RESULT 15
AK014456
LOCUS
DEFINITION
AK014456 3239 bp mRNA linear HTC 05-DEC-2002
Mus musculus 13 days embryo whole body cDNA, RIKEN full-length
enriched library, clone:3930401K13 product:CYTOKINE-LIKE NUCLEAR
FACTOR N-PAC homolog [Homo sapiens], full insert sequence.
ACCESSION AK014456
VERSION AK014456.1 GI:12852319
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE
2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE
3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
```


REFERENCE AUTHORS	4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staubli, P., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyo-oka, K., Wang, K. H., Weitz, C., Wittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S., and Hayashizaki, Y.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409 (6821), 685-690 (2001)
MEDLINE	21085660
PUBMED	11217851
REFERENCE AUTHORS	5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
REFERENCE AUTHORS	6 (bases 1 to 3239) Arakawa, T., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT	Please visit our web site (http://genome.gsc.riken.go.jp/) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5', GAGAGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTTTCTTN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 60.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATTCGAGTTAATTAATCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.
FEATURES	Location/Qualifiers 1..3239 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J"

db_xref="FANTOM DB:3930401K13"	
db_xref="MGI:1907072"	
db_xref="taxon:10090"	
clone="3930401K13"	
tissue type="whole body"	
clone_lib="RIKEN full-length enriched mouse cDNA library"	
dev stage="13 days embryo"	
8..1648	
/note="unnamed protein product; CYTOKINE-LIKE NUCLEAR FACTOR N-PAC homolog [Homo sapiens] (SPTR Q9BXK2, evidence: FASTY, 97.8%ID, 100%length, match=1638) putative"	
/codon start=1	
/protein id="BAB29363.1"	
db_xref="GI:12852320"	
db_xref="MGI:1921272"	
/translation="MAAVSLRLGLVWGLGRYPWPVKIVNPPKDLKPRGKKCFV KFFGTKNHAWIKVEQLKPYHAHKEEMIKINKGRFOQAVDAVEEFLRAKGDQTSSTH TSADKDWNSSEERSRPNSGDEKRLSLSEGVKNMGEGKKRVTSSSADRGSKCLK RAQEOSPRKRGRPPKDEKDLTIPESTVKGMMAGPMAAFKWQPTATEPVKADPHFHH FLLSQTEKPAVCYQAITKKLKI CEETGTSIQADSTAVNGSITPTDKKIGFLGLG MGSGIVSNLLKMGHTVWNRTPAEKRGARLGRTPAEVVTCDITFACVSDPKAAKDLV LGPSGVLGIRPGKCYVDMSTVDATVTELAQVIVSRGGRFLEAPVSGNQLSNDGML VILAAAGDRGLYEDSCSFQAMGKTSFFLGEVGNAAKMLIVNMVQSFMTIAEGLTL AQVTGSSQQTLLDILNQQLASIFLDQKQNILQGNFKPDFLYLKYIKDLRLAIALGD AVNHPTPMAAAANEVYKRAKALDQSDNDSMAVYRAYIH"	
BASE COUNT	832 a 785 c 837 g 785 t
ORIGIN	
Query Match	84.4%; Score 630.2; DB 11; Length 3239;
Best Local Similarity	92.3%; Pred. No. 5.1e-155;
Matches	675; Conservative 0; Mismatches 53; Indels 3; Gaps 1;
QY	1 ATGCGGGCTGTGAGTCTGCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCCGATATCCT 60
DB	8 ATGCGGGCTGTGAGTCTGCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCCGATATCCT 67
QY	61 CCTTGGCCAGGAAGATTGTTAATCCACCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAA 120
DB	68 CCTTGGCCAGGAAGATTGTTAATCCACCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAA 127
QY	121 TGCTTCTTTGTGAAATTTTGTGGAACAGAAAGATCATGCTCGATCAAAAGTGGAAACAGCTG 180
DB	128 TGCTTCTTTGTGAAATTTTGTGGAACAAATAATCATGCTCGATCAAAAGTGGAAACACTA 187
QY	181 AAGCCATATCATGCTCATAAAGAGGAAATGATAAAAAATTAACAGGGTAAACGATTCCAG 240
DB	188 AAGCCTTACCATGCTCACAAGGAGGAGATGATAAAGATTAAACAGGGTAAACGGTTCAG 247
QY	241 CAACGGGTAGATGCTGTGGAAGAGTTCTCTCAGGAGAGCCAAAGGAAAGACCAGACGTCA 300
DB	248 CAAGCTGTGGATGCTGTTGAAGAGTTCTCTCAGGAGAGCCAAAGGAAAGACCAGACATCA 307
QY	301 TCCACAAATTTCTTGATGACAAGAAATCGACGTAATTCAGTGAGGAGAGAGTAGGCCCA 360
DB	308 TCCACACTTCTGCTGATGACAAGGATTGGCGTAATTCAGTGAGGAGAGAGTAGGCCCA 367
QY	361 AACTCAGGTGATGAGAAGCGCAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAGACATG 420
DB	368 AACTCAGGTGATGAGAAACCGCAAGCTTAGCCTGTCTGAAGGGAAGGTGAAGAAGACATG 427
QY	421 GGAGAAAGGAAAGAGAGGTGTCTTTCAGGCTTTCAGAGAGAGGCTCCAAATCCCTCTG 480
DB	428 GGAGAAAGGAAAGAGAGGTGACTTCAGGCTTCGACAGACAGAGGCTCCAAATGCTT 484
QY	481 AAAAGAGCCCAAGAGCAAAAGTCCCGGGAAGCGGGGTGGCCCCCAAGGATGAGAAGGAT 540
DB	485 AAAAGAGCCCAAGAGCAAAAGTCCCGGGAAGCGGGGTGGCCCCCAAGGATGAGAAGGAC 544
QY	541 CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGGAACCGATGGCCCGTTT 600
DB	545 CTCACCATCCCTGAGTCTAGCACTGTAAAGGGGATGATGGCTGGACCGATGGCTGCATTT 604

THIS PAGE BLANK (USPTO)